With the compliments of
The authors

MENTAL TESTS AND HEREDITY

INCLUDING A SURVEY OF NON-VERBAL TESTS

BY

BARBARA SCHIEFFELIN
AND
GLADYS C. SCHWESINGER

1930

PUBLISHED BY
THE GALTON PUBLISHING COMPANY, INC
11 WEST 42ND STREET, NEW YORK, N. Y.

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

This volume is the outcome of a search for "tests of innate ability". It has developed into a revelation of the difficulties to be encountered when mental tests are used to measure mental endowment.

The work was projected by the Eugenics Research Association in the fall of 1928, as one of a series of studies of the present status of the sciences contributory to the sound development of eugenic research. A committee of the Research Association, composed of Drs. Clark Wissler, Charles B. Davenport, Harry H. Laughlin, Charles G. Campbell and Frederick Osborn outlined the original project, and it was then turned over to the Carnegie Institution of Washington. Under its auspices, a committee was formed on Maintenance and Extension of Research Activities in the Field of Eugenics, consisting of Drs. A. V. Kidder (chairman), Carl C. Brigham, Charles B. Davenport, Leslie C. Dunn, Harry H. Laughlin, Rudolph Pintner, Edward Lee Thorndike, and Clark Wissler. The help and constructive criticism of this committee were of invaluable assistance as the work developed.

During the course of the study, many authorities were consulted by means of personal interviews and correspondence. To these psychologists and anthropologists, who gave of their time and interest (their names appear throughout the book), sincere thanks and appreciation are tendered.

Grateful acknowledgement is also made to the publishers who permitted the use of quotations and illustrations in the text. These include the Cambridge University Press, Columbia University Publications, Ginn and Company, the Harvard Graduate School of Education, Alfred A. Knopf, Little Brown and Company, W. W. Norton and Company, the Princeton University Press, the Psychological Review Company, the Public School Publishing Company, the Science Press, the C. H. Stoelting Company, Warwick and York, Inc., and the Williams and Wilkins Company.

Source material was supplied by many published books and articles. Special indebtedness is due to the following: "A Manual of Individual Tests and Testing" by Bronner, Healy, Lowe and Shimberg, "Mental Tests—their History, Principles and Practice" by F. N. Freeman, "Intelligence Testing" by R. Pintner, "Mental Tests in Clinical Practice" by F. L. Wells, "A Scale of Performance Tests" by Pintner and Paterson, "A Universal Scale of Individual Performance Tests" by P. C. Squires, "International Group Mental Tests" by S. C. Dodd, "A Study of Mental Testing in Relation to Anthropology" by B. Blackwood, "An Investigation into the Validity of Norms, with Special Reference to Rural and Urban Groups" by M. Shimberg, and "The Psychological Register", edited by Carl Murchison.

BARBARA SCHIEFFELIN,
GLADYS C. SCHWESINGER.

New York City,
September 30th, 1930.

TABLE OF CONTENTS

INTRODUCTION

This book has two purposes. The first and more general one is to outline the problems involved when mental tests are used for the purposes of eugenics. The second is to survey as completely as possible the non-verbal tests which have been offered at various times for use in appraising "general intelligence." If eugenic research is to be defined as "the study of the part played by human affairs in heredity", a need arises for tools with which to pursue that study. Changes in the human germ plasm are measurable only in the individual, the phenotype; objective measurements of these final traits, mental and physical, are therefore a* sine qua non *for eugenic research.* (See page 31 of this volume.)

"Human affairs" includes all the non-biological factors that go to make up a "culture" in the anthropological sense. The term "environment", though actually including much more than culture (climate, geography, food supply, etc.) will here be limited to mean "human affairs." The scientific method of dealing with two variables such as heredity and environment is to compare groups of organisms, each group being theoretically the same in one variable and different in the other. When human beings are the subjects of such comparisons, a controlled procedure is complicated by the enormous number of factors within each variable. For instance, if two groups could be obtained differing as to race but with a similar cultural background, a real comparison might be made. Actually, this is never the case, because it is almost impossible not only to isolate pure races, but also to equalize differences in culture. If one of these differences could be eliminated, the task of controling the environmental factor in certain group comparisons would be simplified. It is for this reason that tests with relatively few verbal elements were selected for this survey, even though it is true that freedom from language does not imply freedom from the effects of training or local knowledge. Before any tests can be considered valid tools for the comparison of groups —and hence for the measurement of native differences—it will be necessary for leaders in the field of mental testing to rank them according to their applicability to specific groups which differ in specific ways and then for the highest ranking tests to be actually tried out.

The first part[2] of this study consists of a brief historical sketch of the testing movement and of present-day theories of intelligence. Part Two gives a general discussion of the heredity-environment issue, the underlying theme being that

** See Reference No. 1, in the Appendix (page 259) for an elaboration of this definition. Hereafter, all "footnotes" will be found in their numerical order in Appendix. In order that the text shall not be over-burdened with references, these appendices have been made rather long, especially as regards Part II. This is necessary for two reasons. First, Part II is a bird's eye view of many fields of research, and therefore the references to various workers are frequent; second, the material for the section on inter-racial and universal testing was based on information gained during many personal interviews with leading psychologists and anthropologists. It was thought that the least clumsy way of handling the many citations thus involved was to use the appendix technique.*

hereditary mental propensities can be studied only through their interplay with environmental processes. The various ways in which mental tests have been employed to measure the influence of outer circumstances on the innate endowment of different human groups are examined, and the three commonest techniques for studying mental inheritance are outlined. One of these techniques, namely, the comparison of groups, as exemplified by inter-racial testing, is dealt with in greater detail than the other two. The difficulties inherent in racial studies are summarized, and the tests which have been developed for inter-racial use are described. Representative opinions are given on the possibilities and limitations of "universal" testing, and the problems thus revealed lead to a discussion of a modified method of group comparison for which the name "intra-group scaling" has been coined. This would utilize both the principle of the universal test—i. e. the common measuring unit—and also adhere to the basic testing axiom that the most useful measurements are those best adapted to homogeneous groups. It would require, roughly, the scaling on two specific groups of a test composed of material common to each group but not necessarily "universal". The two resulting sets of items (scaled according to different difficulty values) could then be used on both groups, and the superiority of one group over the other would be computed by comparing the two differences—the smaller one indicating the "best" group. This method of comparison might yield interesting information about differences in environment or heredity between particular groups.

Part Three presents the non-verbal tests which have been collected. Preceding the actual survey is one chapter on mental tests in general and non-verbal tests in particular, and another giving the statistical hazards involved in the interpretation of test results. It must be understood that the information on each test is presented in very general form, and that therefore this survey is not to be regarded as a test handbook. Actually, an inventory such as this, including only one class of tests, merely begins the task of complete collection. A thorough canvass of the testing field for measurements that might be used in isolating innate characteristics would require a sifting out of pertinent data from many divisions of psychological study. Unfortunately the time limits of the survey did not permit of thorough investigation into the following fields:

1. *Tests for infants and pre-school children.*
2. *Objective tests of personality traits, including temperament, character and emotion.*
3. *Special ability tests.*
4. *Sensori-motor tests.*

The first is the most serious omission, because developmental norms for infants come nearer than any other tests to being free from environmental influence. In each of the other categories is also to be found test material of great value. For instance, certain objective gauges of temperament and feeling, such as bodily changes during emotional tension, may well furnish the key to inborn differences of character. Again, tests of "special abilities" such as talent for drawing (which is not necessarily connected with intelligence) are perhaps the most valid tests of individual differences because the performance measured is like the performance in which success is to be predicted. Finally, sensori-motor measurements, though no longer held to be diagnostic of intelligence, still seem to tap something pecul-

iarly characteristic of the individual, which is relatively independent of training. Another phase which could not be included in this survey is the correlation of mental abilities with structural organic bases, such as speed of reflex action, and endocrine development. Whatever inroads the writers have made into the above divisions of mental measurement—especially as regards sensori-motor and special ability tests—will be reported.

It was also found necessary to restrict this survey to non-verbal tests IN AMERICA. *This limitation is to be regretted, for there is undoubtedly much valuable material to be found abroad.*

Finally, this account of the survey's scope should be concluded by emphasizing that it makes no attempt to develop new techniques of measurement, or to evaluate former techniques. It neither deals with the problems involved in the designing or scoring of tests, nor does it seek to throw new light on the interpretation of test results. *(For a detailed outline of the survey, see pages 61–63.)*

The fourth and last section is a general conclusion and summary of results. It gives tentative recommendations as to steps to be taken for the purpose of fostering the development of objective measures of mental traits. The attainment of such generally accepted standardization as would justify the use of the best tests by eugenists is discussed in connection with the question of co-operation between psychology and eugenics. It must be clearly understood, however, that no part of this book deals with such topics as the nature of innate ability, or the causes of differences between human beings. No assumption is made to the effect that mental traits which have never been influenced by environment exist in testable form; indeed, the absurdity of trying to distinguish what fraction of a person's behavior at a given moment is determined by his innate endowment, and what fraction is due to the bombardment of the environment to which he has been subjected, is proclaimed at the risk of verbosity. If one assumes that the abilities which make for successful achievement in a certain environment can be measured by tests composed of materials from a different one, the question of testing innate ability merges into the problem of environmental differences. In the words of Thorndike ('26): ". . . . the problem of analyzing a person's intellectual ability into an amount due to nature or an amount due to nurture, is unsolved. No task or test has been proved to be a measure of the former alone. The wisest procedure at present is to equalize environmental forces by using a wide variety of data with which all individuals have had adequate experience, and to make as correct allowance as we can for what we cannot equalize."* P. 462.

* *A bibliography is appended to each chapter within the main section, where the references in the text and in Appendix I are given in full. The last two numbers of the year in which a work was published are inserted after the author's name, thus enabling the reader to look up the name alphabetically and locate the work chronologically. Thus:*

Thorndike, E. L. (with others) '26. The measurement of intelligence. New York; Teach. Coll. Bur. Publ. Pp. 616.

PART ONE

HISTORICAL SURVEY

Part One, Chapter I

TESTS OF INDIVIDUAL DIFFERENCES

Ever since the teachings of Hippocrates, thoughtful people have recognized that men differed from one another in their perceptive powers, in resistance to fatigue, and in variously defined mental capacities, as well as in their physical traits. Sir Francis Galton was the first to make a scientific study of the inheritance of these "individual differences". The opening sentence in "Hereditary Genius" ('69) runs as follows: "I propose to show in this book that a man's natural abilities are derived by inheritance under exactly the same limitations as are the form and physical features of the whole organic world." As Pintner ('23, p. 18) points out, attention was not focused for long on this inheritance of "natural abilities", or "mental traits", before the necessity arose for methods of estimating their amount. In European psychological laboratories, interest in the study of individual differences was stimulated by experiments aiming to discover the general laws of human behavior. At first, the variations that occurred during the experiments were thought of merely as errors, and, consequently, scientific interest in human differences *per se* did not begin until around 1890. Galton, faced in the sixties with the necessity of discovering some method of measuring individual differences in order to study their heritability, was therefore thrown back on such means as questionnaires and ratings. In "Hereditary Genius", however, he devised an imaginary scale, based on the theory of a normal distribution, for the measurement of general ability. This introduced two new and influential ideas: first, the concept of General Intelligence (importantly developed for the first time by Binet, thirty-five years later), and second, the theory that people cannot be grouped into distinct types, such as idiots and geniuses, because they differ from each other in general ability by measurable amounts. Galton also devised instruments to study differences in sensation, of which the best known is the Galton whistle, constructed to measure the highest tones that can be heard by human ears.

Galton

Associated with Galton was J. McKeen Cattell, who came from Wundt's laboratory, where he had been the first both to measure variations in reaction time and to establish the concept of individual differences. This experience, together with Galton's emphasis on the importance of studying these differences, stimulated Cattell's experimentation with mental measurements, which in turn initiated the testing movement in the United States. In an article entitled "Mental Tests and Measurements" ('90) Cattell first introduced the term "mental test". The tests described in this article measured acuity of sensation, speed of movement, simple judgment and simple memory. Eleven years later, Clark Wissler ('01) reported on the results of similar tests, with the addition of three more complex tests of association, imagery and

Cattell

memory. The scores of these tests were correlated in order to show what relationships existed among the various tests themselves, and between test scores and class grades. The connection between the tests proved to be so small that the author concluded that independent special abilities were being dealt with. Again, the test scores were no more closely related to scholastic ability than they were to each other, while it was shown that class grades had high inter-correlations. Freeman ('26, p. 48) advances reasons to explain these results, chief of which is that the mental processes tested were mainly sensory and motor — functions now generally admitted to be related neither to one another nor to achievement in school or vocational situations. The fact that the tests themselves might not have been reliable measures of the subject's capacity may also explain these low correlations.

Wissler

Other early mental testers in America, Gilbert ('94), Bolton ('91) and E. A. Kirkpatrick ('00), sought to measure the relationship between a subject's score on a test, such as memory span for digits, and some other measure of his powers, usually teachers' estimates of mental ability. Sharp ('99) influenced by Binet, tested some eight students with memory, imagery, imagination and attention tests. Bagley ('01) and Seashore ('99) made studies on the relation between motor and sensory processes and mental ability. On the whole, any correlations found were too small to be of practical importance in diagnosing iintelligence. Because of these negative results, American psychologists lost interest in the invention of mental tests during the succeeding few years; their chief concern was, in the words of Freeman, "the minute standardization of a few tests of simple sensory and motor processes."

Other Early Testers

About this time a practical interest in the problem of abnormal children arose and sociologists began to ask psychologists for tools with which they could diagnose amentia. In 1903 R. L. Kelly's article on "Psycho-physical Tests of Normal and Abnormal Children" appeared, reporting a general increase in motor co-ordination with an increase in intelligence. G. E. Johnson ('95) who compares the performances of feeble-minded and normal children, stresses an urgent need for standards. Norsworthy ('06) was another pioneer in the testing of feebleminded children. Just before the appearance of the Binet scale, she published a report which, by expressing the "variability of the child in terms of the variability of the group" was one of the first examples of "standardization".

Almost the only criteria for the tests used in these early studies were preliminary "arm-chair" analyses of the mental processes they were supposed to measure. The first powerful objector to the view that it was desirable to "measure definitely a definite thing" was Alfred Binet. He held that one must test the *complex* processes, for only in them are variations sufficiently great to differentiate human beings. Even though more consistent results are obtained in testing the simple processes, they are not valid indicators of intelligence. Kraepelin, (interested primarily in diagnosing insanity), Ferrari, Oehrn, Jastrow and Ebbinghaus had found that complex functions were measureable, but this discovery was made only after having encountered enormous difficulties, some of which are still unresolved. Specific processes were found to vary with the attitude or physiological condition of the subject so that Binet

Binet

finally rejected specific tests of such processes as memory, mental imagery, attention, and the like, because he did not believe that they tested the assumed functions. He concluded that one cannot say of an individual "He has a good memory" or "He is mentally quick"; one must specify in what mental *performances* he is quick or slow, or of which *kind* of memory we speak,— whether it is for digits or letters arranged in a certain way, abstract things or concrete things. [See Peterson ('26)]. Binet and his co-worker Simon ('05) finally decided that, since it was impossible to disentangle the various intellectual processes, their combined functional capacity must be tested, with no pretense at measuring the exact contribution of each to the total product. Terman ('16) concludes an account of Binet's work with the statement: "It is hardly too much to say that intelligence tests have been successful just to the extent to which they have been guided by this principle."

Binet's contributions are too well known and have been described too often to necessitate full treatment here. Mere mention must be made of his work with feebleminded and bright children ('99), his concept of "general intelligence", his brilliant recognition of the value of common experience as a basis for intelligence tests, and, with Simon, his pioneer work in developing a mental age scale with yearly increments ('08).

Later Developments

The history of the introduction of the Binet Scale into America, and of its various revisions, resulting from the controversy that ensued over its validity, is also far too long and complex to be discussed here. [See Bobertag ('09), Decroly and Degand ('10), Binet ('11), Goddard ('11), Kuhlman ('12), and Terman *et al* ('12 and '16)]. Two studies should be singled out for mention, however— those of Ayres ('11) and Strong ('13) — because they were among the first to note that some of the Binet items *did* depend on "recent environmental experience".

Three other important developments in the history of tests must be noted in passing: (1) Spearman's psychology of cognition, supplemented by certain statistical devices which he invented; (2) the group tests constructed to answer the demand for psychological measurements in the United States Army; and (3) the spread of the testing movement in educational and clinical psychology. For adequate accounts of these, one must turn to special articles or historical surveys. (See K. Young ('24), Pintner ('23), Part I, and Freeman ('26), Chapters 6 and 7).

Sub-tests In Group Scales

Present-day group tests (a "group" as against an "individual" test being one which can be given to more than a single subject at the same time) are composed of various sub-tests. Chief among these is the test known as "Completion of Complex Forms", which Ebbinghaus devised in 1879. This test, according to Terman and Childs ('12) "requires the subject to relate given fragments" (usually of a sentence; sometimes of pictures) "into a meaningful whole". It is so highly thought of that Colvin ('22) calls it "the most important intelligence test contributed by psychologists for determining individual differences". A second important type of test is the analogy. Here the examinee is given a series of three words or pictures to which he must add a fourth bearing the same relation to the third as the second does to the first. Analogies can be classified as associations of the part with the whole (hand : arm), genus with species (mammal : whale), cause with effect (fire : burnt), and similarities with differences (friend : enemy). "Absurdities" is the name given to another kind of test, which requires the subject to isolate irrational elements in sentences or pictures. Perhaps a better known form of test is the maze, which has been used to test everything from feeblemindedness to "social intelligence". This is usually a pencil and paper labyrinth which must be traced through to a goal, but there also exist many three-dimensional mazes. Another widely used test is the Substitution Test; this involves learning to connect one set of characters with another, (usually numbers with letters, or symbols with numbers) according to a key at the top of the page. Cube Analysis and Number Series Completion are two sub-tests which are often included in group scales. The first calls for visual imagery, by having the subject count mentally the exact number of cubes in a picture or model of piled cubes. The second requires the subject to perceive a rhythmic principle underlying a set of numbers (2 - 4 - 6), and to continue the progression. Other tests which presuppose schooling and knowledge of language are: vocabulary tests, directions, arithmetical problems, reading tests, interpretation of proverbs and general information. (Rote memory and "cancellation" tests were formerly much employed, but now play an inconspicuous role in group tests because they do not tap abilities needed in daily life.)

We have observed that while the early interest in tests for specific capacities was waning, a new conception of mental tests as measurements of something called "general intelligence" arose. Around the time of Binet's last revision (1911) and reverting back to the early interest in specific tests, appeared some new and entirely different tests which were thought to measure *specific* mental traits, such as planning capacity, form perception and motor ability, as against general intelligence. They were developed in answer to two needs: First, there were certain cases, such as deaf or foreign-born subjects, for whom language difficulty or culture differences made the Binet test unfair. Second, the existing so-called general intelligence tests were thought to measure only the more abstract part of an individual's mental capacity. (See Freeman ('26), p. 106.) Consequently, in order to make a more universal and all-round inventory, a new kind of test was invented which was composed, for the most part, of non-language problems in the shape of manual performances, such as formboard fittings, puzzle-picture completings and cube tapping imitations, or problems requiring for solution some simple act of drawing, similar to the drawing items in the Binet Scale (e.g. mazes, cube analysis, digit-symbol and number checking). Many of these tests were organized into scales in which the different tests did not necessarily correlate with one another, and the subject's reactions to the various items were kept distinct. This, according to Freeman ('26, p. 108), was done in order to discover prominent traits, and also to determine the subject's general mental level by a summary discussion of responses to single tests, instead of by the calculation of a composite numerical score. The best known of these test groups are the Healy-Fernald tests ('11), the Pintner-Paterson Performance Scale ('17), the Army Beta (Yerkes '21), the Porteus Maze Scale ('15), the Woolley Scale ('15) and the various graded series of formboards: Ferguson ('20), Dearborn *et al*

Performance Tests

('23), Kent ('28), and the like. Others, less widely used, are Squire's Graded Mental Tests ('12), Knox's Scale for Testing Immigrants ('14), Mullan's Scale of 100 Points for Testing Immigrants ('17) and Habermann's Intelligence Examination ('16). These different scales were composed chiefly of non-language tests, each of which was supposed to tap definite mental capacities. Many performance tests exist, of course, outside of the scales, such as the Witmer Formboard ('11) and the Kohs Block Design Test ('23). New performance, non-language and special ability tests have been appearing in numbers ever since the Pintner-Paterson Scale. The Whipple Manual gives comprehensive information on individual tests prior to 1919; the best account of performance tests up to 1927 may be found in the Bronner-Healy "Manual of Individual Tests".

It is, of course, a moot question whether the principle underlying performance tests is fallacious or not—i. e. whether the previous experience on which they are based is *more or less specialized* than that assumed by verbal tests. (For a discussion of this principle, see Part III, Chapter I.)

At present, the mental testing movement can be said to be slowly recovering from the false impetus it received from the misinterpretations of the Army results. The most hopeful signs for the future are a constantly increasing caution in the construction and validation of new tests and in the interpretation of scores, coupled with a general lack of confidence in a standardized test simply because it is standardized. Before progress can be made in intelligence testing, intelligence theory must be based on sound experiment.

BIBLIOGRAPHY

References cited in Part I, Chapter I.

Tests of Individual Differences.

Ayres, L. P. '11. The Binet-Simon measuring scale for intelligence. Some criticisms and suggestions. Psychol. Clin., 5, 187–196.

Bagley, W. C. '01. Mental and motor ability. Amer. J. Psychol., 12, 193–205.

Binet, A. '99. Attention et adaptation. Année Psychol., 6, 248–404.

—'11. Nouvelles researches sur la mésure du niveau intellectuel che les enfants d'école. Année Pspchol., 17, 145–201.

— and Simon, T. '05. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. Année Psychol. 11, 191–244.

— —'08. Le developpement de l'intelligence chez les enfants. Année Psychol. 14, 1–90.

Bobertag, O. '09. A. Binet's Arbeiten uber die intellectuelle Entwicklung des Schulkindes. Zeitsch. fur Angew. Psychol. B. 3, 230–259.

Bolton, T. L. '91. The growth of memory in school children. Amer. J. Psychol. 4, 362–308.

Bronner, A. F., Healy, W., Lowe, G. M., and Shimberg, M. E. '27. A manual of individual mental tests and testing. Boston: Little Brown and Co. Pp. 287.

Carr–Saunders, A. M. '26. Eugenics. New York: Henry Holt and Co. Pp. 252.

Cattell, J. McK. '90. Mental tests and measurements. Mind, 15. 373–380.

Colvin, S. S. '22. Principles underlying the construction and use of intelligence tests. 21st Yrbk. Natl. Soc. Stud. Educ. 11–45.

Dearborn, W. F., Shaw, E. A. and Lincoln, E. A. '23. A series of formboard and performance tests of intelligence. Harvard Monog. Educ., 1, No. 4, Pp. 64.

Decroly, O. and Degand, J. '10. La mésure de l'intelligence chez les enfants normaux d'apres les tests de Binet et Simon. Arch. de Psychol., 9, 81–108.

Ebbinghaus, H. '97. Uber eine neue Methode zur prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. Zeitsch. f. Psychol. B. 13, 401–459.

Ferguson, G. O. '20. A series of formboards. J. Exp. Psychol. 3, 47–58.

Freeman, F. N. '26. Mental tests—their history, principles and applications. Boston: Houghton, Mifflin Co. Pp. 503.

Galton, F. '69. Hereditary genius. An inquiry into its laws and consequences. London and New York: MacMillan Co., 2nd Ed. 1892, Pp. 379.

Gilbert, J. A. '94 Researches on the mental and physical development of school children. Studies from the Yale Psychological Laboratory, 2, 40–100.

Goddard, H. H. '11. A revision of the Binet scale. Vineland Training School Bull. 8, 56–62.

Haberman, V. '16. The intelligence examination and evaluation. Psychol. Rev., 23, 352–379, 484–500.

Healy, W. and Fernald, G. '11. Tests for practical mental classification. Psychol. Monog. 13, Pp. 54.

Johnson, G. E. '95. Contribution to the psychology and pedagogy of feebleminded children. Ped. Sem. 3, 246–301.

Kelley, R. L. '03. Studies from the psychological laboratory of the University of Chicago. Psychol. Rev. 10, 274–380.

Kent, G. H. '28. A graded series of formboards. Person. J., 7, 115–120.

Kirkpatrick, E. A. '00. Individual tests of school children. Psychol. Rev. 7, 274–280.

Knox, H. '14. A scale based on work at Ellis Island for estimating mental defects. J. Amer. Med. Assoc. 62, 741–747.

Kohs, S. C. '23. Intelligence measurements. New York: MacMillan Co. Pp. 312.

Kuhlman, F. '12. A revision of the Binet Simon system for measuring the intelligence of children. J. Psycho-Asthenics, Monog. Suppl. No. 1, Pp. 41.

Mullan, E. H. '17. The mentality of the arriving immigrant. Public Health Bull. 90, U. S. P. H. Service.

Norsworthy, N. '06. The psychology of mentally deficient children. Archives of Psychol. 1. Pp. 111.

Peterson, J. '26. What intelligence tests are based on. Indust. Psychol. 1, 569–579.

Pintner, R. '23. Intelligence testing. New York; Henry Holt and Co. Pp. 406.

— and Paterson, D. G. '17. A scale of performance tests. New York: Appleton Co. Pp. 218.

Porteus, S. D. '15. Mental tests for feebleminded; a new series. J. Psycho-Asthenics, 19, 200–213.

Seashore, P. E. '99. Some psychological statistics. Univ. of Iowa Stud. in Psychol. 2.

Sharp, S. A. '99. Individual psychology. A study in psychological method. Amer. J. Psychol. 10, 329–391.

Spearman, C. '04. General intelligence objectively determined and measured. Amer. J. Psych. 15. 201–292.

Squires, C. R. '12. Graded mental tests. J. Ed. Psychol. 3, 363–380, 430–443, 493–506.

Strong, A. C. '13. 350 white and colored children measured by the Binet Simon measuring scale of intelligence. Ped. Sem. 20, 485–513.

Termon, L. M. '16. The measurement of intelligence; an explanation of and a complete guide for the use of the Stanford revision and extension of the Binet–Simon intelligence scale. Boston: Houghton Mifflin Co. Pp. 362.

— and Childs, H. G. '12. A tentative revision and extension of the Binet–Simon measuring scale of intelligence. J. Ed. Psychol. 3, 61–74; 133–143; 198–208; 277–289.

Whipple, G. M. Manual of mental and physical tests. Baltimore: Warwick and York.

1st ed. 1910 Pp. 534.

2nd rev. ed. Part I. 1914 Pp. 354.
Part II. 1915 Pp. 336.

3rd ed. with added references:
Part I. 1924 Pp. 367.
Part II. 1925 Pp. 349.

Wissler, C. '01. The correlation of mental and physical traits. Psychol. Rev. 3, 1–63.

Witmer, L. '11. Courses in psychology at the summer school of the University of Pennsylvania. Psychol. Clinic 4, 245–273.

Woolley, H. T. '15. A new scale of mental and physical measurements for adolescents. J. Ed. Psychol. 6, 521–550.

Yerkes, R. M. '21. Psychological examining in the United States army. Mem. Nat. Acad. Sci. 15, Pp. 890.

Young, K. '24. The history of mental testing. Ped. Sem. 31, 1–48.

Part One, Chapter II

THEORIES OF INTELLIGENCE

Closely associated with the historical development of measures for mental traits, and in many cases originating from newly invented testing methods, are the modern theories of intelligence. To give the full historical background of these theories would be to go back to the earliest Greek philosophers; here we can only view the hypotheses stimulated by present day psychology.

The eugenist who wishes to measure native ability must be in touch with intelligence theory because the question as to whether intelligence is general or specialized has a vital bearing on the principles of heredity. The fact that mental tests have proved useful in the measurement of "general intelligence" does not mean that they can accurately measure "mental traits", which are the units through which the laws of heredity must operate. In the words of Witty and Lehman ('30):

> "The very definition of general intelligence (as a sum total of closely related abilities) precludes the possibility of parallelling the inheritance of physical traits until each of the mental abilities has been identified, and its inheritance demonstrated empirically." (See also page 31 of this volume.)

As has often been admitted, test makers have continued to construct mental tests even though the true nature of intelligence is still a matter of speculation. Rather than depreciating mental tests, however, this admission points to their essential soundness. The modern psychologist utilizes the every-day, commonsense judgment of the world as the criterion of ability. According to Pintner ('23), Binet had to discover what reactions were common to different subjects who had previously been judged "intelligent" or "unintelligent" by their parents or teachers, before the intelligence of unknown subjects could be judged by test criteria. Pintner further indicates that the difference between intelligence and knowledge, so hazy in popular thought, was recognized by Binet only after much experience with his scale of tests. Today, the accurate diagnosis of what the tests test is being undertaken by the process known as "item analysis". (See Appendix I, No. 48) Indeed, the effort to discover which tests are most immune from the effects of environment is a fundamental form of "item analysis".

Before describing the chief present day theories of intelligence, it might be well to give an indication of the number and variety of short *definitions* which have been reached by different authorities. Lest the reader consider some of them too arbitrary, let it be remembered that most psychologists agree in thinking that interpretation of a subject's behavior in a test situation must be more or less speculative. The most one can say accurately is that the subject is good or poor in the specific situations presented by the separate parts of the test,[3] and even then, factors such as the background of the subject, his health and emotional state, must be carefully considered.

Definitions

A few of the more outstanding definitions of intelligence follow:[4]

1. L. W. STERN's definition: ('14)

"General intelligence is the ability of the organism to adjust itself adequately to new situations."

This is interpreted by Pintner ('23) as meaning, essentially, ease in breaking old habits and in forming new ones, which goes back to the general modifiability of the nervous system.

2. A. BINET's definition: (with Simon, '16)

"Intelligence is the sum total of all those thought processes which consist in mental adaptation."

The three fundamental thought processes are, according to him: (1) The tendency to take and maintain a definite direction. (2) The capacity to make adaptation for the purpose of attaining a desired end. (3) The power of auto-criticism. Various other writings of Binet's emphasize attention, adaptation, judgment and initiative as important factors in general intelligence.

3. R. S. WOODWORTH's definition: ('21)

The important factors in intelligence are: (1) The ability to learn and remember. (2) Responsiveness to relationships—i. e. the ability to seize the key to a novel situation, and ability to adapt oneself to it. (3) Curiosity. (4) Firmness in limiting activity to the right direction and suppressing useless acts.

4. C. BURT's definition: ('09)

"Intelligence is the power of readjustment to relatively novel situations by organizing new psycho-physical combinations."

He speaks of it further as the "capacity for continually systematizing mental behavior by forming new psycho-physical co-ordinations, old co-ordinations being retained, so that new co-ordinations bring with them increased complexity and incessant change .... In such progressively integrative actions of the mind, the efficient and directing agent is attentive consciousness. Voluntary attention is the essential factor of general intelligence."

5. L. WITMER's definition: ('22)

"General intelligence is the ability to succeed in the competitive events of a life-time. It is the ability to solve new problems. It is knowledge applied."[5]

6. S. S. COLVIN's definition: ('22)

"Intelligence is a group of innate abilities by virtue of which the individual is capable of learning in a greater or less degree in terms of the amount of these innate capacities with which he is endowed."

7. L. L. THURSTONE's definition: ('24)

Intelligence is: (a) The capacity to inhibit an instinctive adjustment. (b) The capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trial and error. (c) The volitional capacity to realize the modified instinctive adjustment into overt behavior to the advantage of the individual as a social animal.

8. E. MEUMANN's definition: ('13)

"Intelligence is: (a) The power of independent and creative elaboration of new products out of the material given by the memory and the senses. (b) The ability to avoid errors, surmount difficulties and adjust to the environment."

9. J. PETERSON's definition: ('24)

"Intelligence seems to be a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior."

10. F. N. FREEMAN's definition: ('26)

"Degrees of intelligence are determined by the general capacity of the psycho-physical organism for the formation of new patterns among the elements of experience."

11. F. L. WELL's definition: ('17)

"Intelligence means precisely the property of so recombining our behavior patterns as to act better in novel situations."

12. C. KIRKPATRICK's definition: ('26)
"Intelligence is the innate capacity fixed with the conjugation of the chromosomes in the original cell, which independently of environment determines the differences and resemblances between individuals in ability to learn, the factors of power, speed, accuracy and versatility all being taken into account."

13. P. B. BALLARD's definition: ('22)
"Intelligence is the relative general efficiency of minds measured under similar conditions of knowledge, interest and habituation."

14. H. H. YOUNG's definition: ('16)
"Intelligence is the ability of the individual to solve what for him is a new problem."

15. R. PINTNER's definition: ('26)
"Intelligence, viewed empirically, is a judgment on the part of someone with reference to a specific response made by an individual. The criterion is whether or not the individual is attaining the end he is striving for. There are no specific reactions which are always indicators of intelligence; any reactions may be used, provided the background is the same."

16. S. DODD's definition: ('26)
". . . . for a working concept, intelligence can be thought of as ability to achieve after comparable training." (See Appendix I, No. 55)

Of the many brief generalizations about intelligence, perhaps the most important are the following: Thorndike's ('21) "We may . . . define intellect in general as the power of good responses from the point of view of truth or fact"; Spearman's ('23) concept of intelligence as the "eduction of relations and correlates"; Terman's ('21) "An individual is intelligent in proportion as he is able to carry on abstract thinking" (see also Ebbinghaus '13); Woodrow's ('21) "Intelligence is acquiring capacity"; and Henmon's ('21) "Intelligence is intellect plus knowledge."[6]

Discussion of Definitions

It is evident that these definitions are framed primarily from the standpoint of the mental tester and not from that of the evolutionist. Instead of being couched in terms of the "difference between the amoeba and the intellectual giant", they deal with the general growth process of the human being. There is so much overlapping in these concepts of intelligence, that the following comments would be generally applicable to them: First, as Freeman ('26) remarks, most of them are framed with a view to throwing light on individual differences and their quantitative measurement, rather than to describing the processes and functions of intellect.[7] Second, it is noticeable that the majority of the definitions are couched in terms of behavior. Pieron ('26) says: "Intelligence does not exist in the mental mechanism; it is only . . . . a behavior value. We might as well try to find out the speed of an automobile by taking it apart" (as to try to get a theoretic formula for intelligence). Third, there is a general tendency to trace intelligent behavior back to the neural mechanisms which underly it. Indeed, the definitions of Burt, Peterson, Freeman, Wells and Kirkpatrick verge toward such physiological accounts as Dearborn's "integrations of conditioned response" ('21) and McCall's analysis of intelligence as "a number of easily formed and broken neural connections . . . . with the desirable connections in a state of permanence" ('22). (Of great interest in this connection is the work done by Lashley ('29) on the brains of animals.) Finally, the last four definitions could be singled out as one group in that they are statements of what happens when a person's score in a mental test is taken as indicative of his intelligence.

Another attitude toward intelligence which it seems significant to comment upon, (because of its relation to the problem of measuring innate ability) is one which differentiates between "intelligence" and "intellect". William James drew this distinction,[7] and it has been succinctly expressed by Henmon, quoted above. Here intellect is taken to mean, in the words of Baldwin ('01–05) "the processes of elaboration which the mind applies to its material, and in particular that kind of elaboration which leads to conceptual thinking." At Witmer's Psychological Clinic, intelligence is distinguished from intellect on the basis of the materials worked with. For example, a savage faced with a situation new to him may show intelligence in his behavior, and Einstein tackling a new problem of relativity will also exhibit intelligence; but Einstein will be functioning on a far different intellectual level than the savage.[8]

This distinction between two intellectual levels brings to mind the familiar discrimination between "idea thinkers" and "thing thinkers". Indeed, intelligence has actually been arranged into three qualitative "types"— verbal, concrete and social—according to the kind of stimuli used as test situations. (Pintner '26) Bridges ('26) has elaborated a similar grouping as follows:

1. Cognitive intelligence, or ability to acquire ideational associations.
2. Conative intelligence, or ability to co-ordinate motor responses into habits, and acquire technical skills.
3. Affective intelligence, or ability to condition, modify and combine the feelings.

(Note: Two other less widely accepted "kinds" of intelligence have been postulated: Rugg's "aesthetic intelligence" and Toops' "clerical intelligence".)

Thorndike also distinguishes between abstract, mechanical and social intelligence, and many workers in schools and clinics use this classification for practical testing purposes. Witty and Lehman ('30) criticize it on the grounds that none of the "types" have been "precisely defined and accurately measured". They point out that if this classification really divides intelligence into separate "abilities", then the general intelligence test has been misnamed. In their words, modern tests "measure numerous somewhat related and some unrelated abilities and attainments, which appear to have little or no relationship to the ability to adjust one's self adequately to new situations". Terman ('21) argues for a hierarchy of intellectual qualities[9] because intelligence must be considered from the standpoint of complexity levels rather than of differences in "kind". According to him, conceptual thinking is "higher" than manipulative skill. Certain it is that tests for "general intelligence" involving language — tests of "selective, rational, generalizing and organizing abilities" (Thorndike '26) — are the most useful as prognoses of success in a standardized situation like a school career; that vocabulary proficiency correlates more highly with our criteria of intelligence than does performance ability; and that the complex processes that develop with age are better indicators of intelligence than are the simple sensori-motor functions. (Freeman '26) Able defenses of concrete and social intelligence exist,[10] and there is no doubt that as tests of motor performance and will-temperament become more accurate, new light will be thrown on these processes, and their contribution to "general ability" will be more clearly recognized. (See also page 16 and Part Three, Chapter One).

These definitions and classifications, whether they view intelligence qualitatively or quantitatively, do little more than express the practical views of different psychologists as to what distinguishes bright people from dull. They do not seek to explain the nature of intelligence and its relation to the different traits of the individual; in fact, according to Dr. Carl Brigham", most of them are teleological. There are, however, two outstanding *theories* of intelligence, which are more or less based on experimental evidence. They are, generally speaking, opposed to each other as to whether intelligence is made up of distinct elements more or less interrelated, or whether it is a combination of general ability and of unrelated specialized abilities. These theories are of importance to the mental tester because each one advocates a different way of measuring individual differences.

Theories of Intelligence

The earliest, called the Synthetic Theory, was originated by Wundt and developed by the British Association School. It holds that intelligence is a complex of "associated mental content"; in other words, is the totality of one's several elemental abilities. According to later views, undue stress is laid on the sensory and elementary reactions; for instance, such "faculties" as sensation and perception, now thought of as "unit activities of the total mind", were then considered "separate elements". To test intelligence according to this theory would be to aattach a weight to every trait in proportion as it contributes to the total intelligence, and then add. Stockton ('22) cites Rossolimo's Profile Method, Netschajeff's Tachistoscope and the Yerkes Bridges Point Scale as examples of mental measurements built according to this principle. The difficulty lies, of course, in determining what each trait contributes.

Synthetic Theory

Thorndike has been called a representative of both the first and the second schools. His view is, roughly, that a number of specific abilities exist which are highly inter-correlated because of the elements they possess in common. Where there is a low correlation, there are few common elements; for instance:

> "The efficiency of a man's equipment for the specifically human task of managing ideas is only loosely correlated with the efficiency of the simpler sensori-motor apparatus which he possesses in common with other species" (—et al '09).

and later:

> "The mind must be regarded not as a functional unit, nor even as a collection of a few general faculties which work irrespective of particular material, but rather as a multitude of functions, each of which involves content as well as form, and is related closely to only a few of its fellows, and to the others with greater and greater degrees of remoteness." ('14).

Two Factor Theory

The second theory, postulating Spearman's "g" factor,, resulted directly from the use of the correlation technique in analysing test scores. Its origin can be traced to the fact that mental processes were found to be associated together in greater or less degree. When the army psychologists sought for tests which correlated slightly with each other, and highly with outside criteria, they found that when the latter was the case, the tests correlated highly with each other. All this pointed to the general unity of intelligence, and Spearman's "Two Factor

Theory" has worked out this concept more systematically than any other. It postulates, briefly, that the ability behind a test performance consists of a factor specific to itself, and of part of a "general intelligence factor" or "central fund of energy", "g", which is so homogeneous as to be measurable in any one of the separate functions. Thus, a person's success in memorizing numbers, would depend jointly on his specific memorizing ability and on his general intelligence. Spearman ('25) has even brought forward a physiological explanation of his theory in terms of cortical activity. Freeman ('26) cites this as Spearman's quantitative principle, and quotes him thus:

> "'G' measures something of the nature of an 'energy' derived from the whole cortex or wider area of the brain. Correspondingly, the s's (specific factors) measure the respective efficiences of the different parts of the brain in which this energy can be concentrated; they are, so to speak, its 'engines'. Whenever the mind turns from one operation to another the energy is switched off from one engine to another, much as the power supply of a factory can be directed, at one moment to turning a wheel, at the next to heating a furnace, and then to blowing a whistle."— p. 488 (From Spearman '25)

Interest in the theory of "g" and its implications has never flagged. Many investigators find it adequately supported by their facts, while others attack it. One of its ablest critics is Truman L. Kelley ('28), who considers that "g" is not the same throughout, being sometimes verbal, sometimes spatial or numerical, sometimes a "speed" and sometimes a "memory" factor. Certain other workers have also found "groups" of abilities, such as comprehension, judgment, intellect with words, numbers or symbols, which are made up of primary functions held together by "bonds". These "group factors" cannot all be measured in the same dimension; some, for instance, must be tested for strength, others for speed, others for range, etc. Brigham ('30) states that if independent groups of traits exist, then the scores of tests in which verbal and mathematical materials, for instance, are indiscriminately mingled, cannot be valid. (See Kelley's work with group factors —'27).

Many psychologists are now endeavoring to find what relationship exists between different sub-tests. It has been discovered, for example, that the "bonds" between verbal and mechanical tests are very weak, and again, that a precise memory for digits is less closely related to "abstract" intelligence than is a good vocabulary. Most authorities hold that group factors function side by side with "general" factors, such as maturity or speed.[12] Freeman ('26) concludes his chapter on the application of the correlation method to mental testing by saying:

> "If it proves possible to test special capacities it will be necessary to conform to the facts of intercorrelation in determining where the dividing lines between them lie. If Spearman's two-factor theory is correct, special capacities do not fall into groups, and we should not expect them to be correlated with each other except in so far as they partake of general capacity."— P. 80.

Conclusion

Our interest in the testing of intelligence is somewhat removed from these theoretical considerations. The question which Clark Wissler asked in 1901 is still foremost in our minds: "What elements of activity contribute to the results of our daily efforts, as well as to the results of tests?" It is probable that tests will never call forth the same abilities as do the conditions of real life, because temperamental factors exhibit themselves differently in a test situation than in daily pursuits. Binet himself saw how difficult it would be to test imponderables, such as richness of inspiration and accuracy of judgment, and many other psychologists have admitted that traits which condition intelligence—interest, perserverance, initiative, imagination, efficiency, insight,— are not subject to quantitative measurement. Personality is what makes or mars a career; and a fine intelligence means little if it is "blocked" emotionally. It is unnecessary to point out that children with high I.Q.'s may lack qualities essential to success, while others who later attain distinction are not discovered by intelligence tests. Stockton ('22) says:

> "There is a type of abstract judgment that is easily missed by any tests . . . . Many of the world's supreme problems have been solved by men who have shown a dogged persistency in pursuing an idea until its relation to other ideas and their relation to it became apparent."

Certain temperaments[13] become flustered when hurried, and are thus set off to poor advantage by tests scored for speed. A boy who shows remarkable ingenuity in fixing the plumbing or rigging up a night lamp (both situations new to him) might conceivably have a hard time with the Healy Puzzle Box because of embarrassment or nervousness at being tested. Despite all this, however, psychologists still think that the scores of intelligence tests furnish useful evidence concerning the intelligence of human beings.

What we are interested in is the fact that nature, in the words of the Indian psychologist, R. N. R. Sarma, is "guilty of partiality in endowing her children with abilities and opportunities". We are concerned with human individual differences in the various "intelligences", and in sensory, muscular, glandular and nervous constitutions. We know that behavior patterns cannot function apart from their environment, just as the stomach cannot digest without food. To discover how actual accomplishment changes with varying conditions, or in other words, in what manner achievement is related to native endowment when the influence of circumstance is as far as possible accounted for, is our quest.

BIBLIOGRAPHY

References cited in Part I, Chapter II.

Theories of Intelligence

Baldwin, J. M. '01–05. Dictionary of philosophy and psychology. New York: MacMillian Co. 3 v. in 4.
Ballard, P. M. '22. Group tests of intelligence. London: Hodder and Stoughton, Pp. 252.
Beauchamp, R. O. and Webb, H. A. '27. Resourcefulness, an unmeasured ability. Sch. Sci. and Math. 27, 457–465.
Binet, A. and Simon, T. '16. The development of intelligence in children. Tr. by E. S. Kite. Vineland Training School. Pp. 336.
Black, C. '28. Note on the nature of intelligence. Brit. J. Psychol. 18, 451–454.
Boring, E. G. '23. Intelligence as the tests test it. The New Republic 35, 35–37.
Bridges, J. W. '26. A theory of personality. J. Abn. and Soc. Psychol. 20, 362–370.
Brigham, C. C. '28. Third annual report, Commission on Scholastic Aptitude Tests, College Entrance Board. New York, Pp. 51.
—'30. Intelligence tests of immigrant groups. Psychol. Rev. 37, 158–165.
Bruce, H. A. '22. Mental tests tested. Cent. 105, 214–221.
Burt, C. '09. Experimental tests of general intelligence. Br. J. Psychol. 3, 94–177.
Carroll, R. P. '28. What is intelligence? Sch. and Soc. 28, 792–793.

Claremont, C. A. '28. Intelligence and mental growth. New York, W. W. Norton and Co., Pp. 120.

Colvin, S. S. '22. Principles underlying the construction and use of intelligence tests. 21st Yrbk. Natl. Soc. Stud. Educ. 11–45.

Dearborn, W. F. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 12, 210–211.

—, Shaw, E. A. and Lincoln, E. A. '23. A series of formboard and performance tests of intelligence. Harvard Monog. Educ. 1, No. 4, Pp. 64.

Dodd, S. C. '26. International group mental tests. Princeton, N. J.: Prince. Univ. Store, Agents. (mimeographed). Pp. 101 plus Appendix.

—'28. The theory of factors. II. Psychol. Rev. 35, 211–279.

Dwelshauvers, G. '28. Le sens du concret et l'intelligence globale (ou facteur G). Psychol. et vie; Rev. de Psychol. appl. 2, 187–190.

Ebbinghaus, H. '13. Grundzüge der Psychologie. Vol. II. Leipzig.

Edwards, A. S. '28. Intelligence as the capacity for variability or versatility of response. Psychol. Rev., 35, 190–219.

Farmer, E. '27. A group factor in sensory motor tests. Brit. J. Psychol. (Gen. Sect.) 17, 327–334.

Freeman, F. N. '26. Mental tests — their history, principles and applications. Boston: Houghton Mifflin Co. Pp. 503.

Gavit, J. P. The definition of intelligence. Sch. and Soc., 29, P. 228.

Henmon, V. A. C. '21. Intelligence and its measurement. A symposium. J. Ed. Psychol, 12, 195–198.

Kelley, T. L. '27. The interpretation of educational measurements, Yonkers, N.Y.: World Book Co. Pp. 363.

—'28. Crossroads in the mind of man. A study of differentiable mental abilities. Stanford University, California: Stanf. Univ. Press, Pp. 238.

Kirkpatrick, C. '26. Intelligence and immigration. Ment. Measur. Monog. 2. Baltimore: Williams and Wilkins. Pp. 127.

Ko, C. S. '27. The nature of intelligence. (Chinese) Edu. Rev., 19, P. 13.

Lashley, K. S. '29. Brain mechanisms and intelligence. Chicago: Univ. of Chicago Press. Pp. 186.

Laycock, S. R. '29. Adaptability to new situations. Baltimore: Warwick and York. Pp. 170.

Mackaye, D. L. '28. The interrelation of emotion and intelligence. Amer. J. Sociol. 34, 451–464.

Mackie, J. '28. The sampling theory as a variant of the two factor theory. J. Educ. Psychol., 19, 614–621.

McCall, W. A. '22. How to measure in education. New York: MacMillan Co. Pp. 410.

Meili, R. '29. Uber einige Schwierigkeiten bei Intelligenz prüfungen. Psychol. Rundschau, 1, 174–179.

Meumann, E. '13. Experimentelle Pädagogik, Vol. II. Leipzig, Pp. 102.

Oates, D. W. '29. The relation of temperament and intelligence to scholastic ability. Forum Educ. 7, 171–185.

Pear, T. H. '22. Remembering and forgetting. London: Methuen. Pp. 242.

Peterson, J. '24. Intelligence conceived of as a mechanism. Psychol. Rev., 31, 281–287.

Pieron, H. '26. The problem of intelligence. Ped. Sem., 33, 50–60.

Pintner, R. '23. Intelligence testing. New York: Henry Holt and Co. Pp. 406.

—'26. An empirical view of intelligence. J. Educ. Psychol., 17, 608–616.

— and Upshall, C. I. '28. Some results of social intelligence tests. Sch. and Soc. 27, 369–370.

Porteus, S. D. and Babcock, M. E. '26. Temperament and race. Boston: Badger Co. Pp. 364.

Pyle, W. H. '28. The I.Q. and the individual. Sch. and Soc. 27, 788–790.

Schneck, M. M. R. '29. The measurement of verbal and numerical abilities. Arch. of Psychol., 107, Pp. 49.

Spearman, C. '23. The nature of "intelligence" and the principles of cognition. London: MacMillan, Pp. 358.

Spearman, C. '25. Some issues in the theory of "G" (including the law of diminishing returns). Proceedings, Br. Assoc. Section J. Southampton.

Stern, L. W. '14. The physiological methods of testing intelligence. Trans. by G. M. Whipple. Educ. Psychol. Monog., 13, Pp. 160.

—'28. Zur Theorie der Intelligenz. Ber. u. d. Kongr. f. exper. Psychol. 10, 170–171.

Stockton, J. L. '22. Definition of intelligence in relation to modern methods of mental measurement. Psychol. Monog., 30, Pp. 118.

Terman, L. M. '21. Intelligence and its measurement. A symposium. J. Ed. Psychol., 12, 127–133.

Thorndike, E. L. and others '09. The relation of accuracy in sensory discrimination to general intelligence. Am. J. Psychol. 20, 364–369.
—'14. Educational psychology. Vol. III. New York: Teachers College, Columbia University. Pp. 408.
—'21. Intelligence and its measurement. A symposium. J. Educ. Psychol., 12, 124–127.
—'24. Measurement of intelligence. The present status. Psychol. Rev., 31, 219–252.
— et al '26. The measurement of intelligence. New York: Columbia Univ. Publ. Pp. 616.
Thurstone, L. L. '21. Intelligence and its measurement: a symposium. J. Ed. Psychol., 12, 201–207.
—'24. The nature of intelligence. New York: Harcourt Brace and Co. Pp. 167.
Wells, F. L. '17. Mental Adjustments. New York: Appleton Co. Pp. 331.
Wissler, C. '01. The correlation of mental and physical traits. Psychol. Monog. Supp. No. 16, Pp. 62.
Witmer, L. '22. Intelligence—a definition. Psychol. Clin. 14, 65–67.
Witty, P. A. and Lehman, H. C. '30. The dogma and biology of human inheritance. Amer. J. Sociol. 35, 548–563.
Woodrow, H. '21. Intelligence and its measurement: a symposium. J. Educ. Psychol., 12, 207–210.
Woodworth, R. S. '21. Psychology, a study of mental life. New York: H. Holt and Co. Pp. 580. (Rev. ed. '29).
Young, H. H. '16. Physical and mental factors involved in the formboard test. Psychol Clin., 10, 149–167.

PART TWO

HEREDITY AND ENVIRONMENT

With special emphasis on the question of racial differences.

Part Two

INTRODUCTION

To approach mental testing from the angle of eugenics is to come to grips with the eternal riddle: which is more important—heredity or environment? This question applied to a particular person is, of course, meaningless, for it is impossible to conceive of an organism apart from its surroundings. Many scientists[14] have shown that the individual is a function of heredity and environment—the product (not the sum) of two factors, just as the area of a rectangle is one hundred per cent dependent on both its altitude and its width. The heredity-environment riddle becomes pertinent, however, when *differences* between individuals or groups are being treated. If we believe, with Herbert Spencer, that life is a continuous adjustment between internal and external relations, the term "native characteristic" comes to mean "a combination of heredity and environment, whose variations are due to changes in the former." (Dodd '26). To weigh justly the amount contributed by each of these factors is a problem which few workers ignore and which many in different professions have attempted to solve. The educator, working with environment, recognizes that differences in original nature determine the extent to which children can profit from training, and hence seeks to arrange conditions so that inherited capacities will be developed to the full. The biologist, working with heredity, knows that organic life is distinguished from inorganic by its power of responding to stimulus with action, and he too must investigate the relations between innate pattern and environment.

If, then, nature does not exist apart from nurture, it follows that studies of mental inheritance must go hand in hand with studies of environment. The problem is to determine whether observed variations are due to differences in innate ability or to differences in environment, and it can only be solved by dealing with specific situations—by stating "what particular hereditary conditions we are comparing with what given range of environmental conditions." (Holmes '21). Dr. Davenport[11] has pointed out that each individual selects his own environment—e. g. two brothers, one with perfect vision and the other slightly astigmatic—may go to the same class in art school, but will not be subjected to the same influences. In other words, unless two people are identical in their biological make-up, perfect equality of situation for experimental purposes is impossible. (See Appendix I, No. 82). Nevertheless, if it can be proved that abilities still range themselves into a normal distribution in spite of the same environment ("same" in Galton's sense—p. 34), and if genealogical resemblances can be shown to persist under dissimilar environments, a verdict in favor of innate ability may be pronounced.

The innate reactions of mankind have been classified as reflexes, instincts, emotions, and capacities (Pintner '29): this study deals only with the last. To speak of measuring capacity as against training is, as we have already seen, to speak without meaning. Training *must* be taken into account, and for this reason, a few samples of studies classed under the general caption "The Analysis of

Environment" are given in the first chapter. The quantitative and qualitative study of environment is still in its infancy, compared with its corollary, the study of mental heredity. Workers in the latter field have developed well defined techniques—the laboratory methods of human and infra-human genetics which do not fall within the scope of this study, and the methods of social research which make use of mental tests, and which will be briefly summarized in Chapter II. One of these methods—racial comparisons—will be treated in detail because studies which compare groups differing as to race provide clear examples of the problems attending group comparisons in general. Tests as tools of comparison—i. e. "universal" tests, and as tools of classification—i. e. "specialized" tests, will be critically discussed, and a new theory advanced as to their use in the comparison of groups.

BIBLIOGRAPHY

References cited in the Introduction to Part II.

Carr, H. A. '25. Psychology. New York: Longmans Green and Co. Pp. 431.

Child, C. M. '24. Physiological foundations of behavior. New York: H. Holt and Co. Pp. 330.

Conklin, E. G. '23. Heredity and environment in the development of men. Princeton, N. J.: Prince. Univ. Press, 5th Ed. Pp. 379.

Darwin, L. '13. Heredity and environment. Eugen. Rev., 5, 153–154.

Davenport, C. B. '11. Heredity in relation to eugenics. New York: H. Holt and Co. Pp. 298.

Dodd, S. C. '26. International group mental tests. Princeton, N. J.: Prince. Univ. Store, Agents. (Mimeographed). Pp. 101 plus Appendix.

Draper, G. '29. Biological philosophy and medicine. Human Biology 1, 117–135.

Dunlap, K. '27. Social psychology. Baltimore: Williams and Wilkins Co. Pp. 261.

Elderton, E. M. '09. The relative strength of nature and nurture. Eugen. Lab. Lect. Series, 3.

Gates, A. I. '28. Observed facts and theoretical concepts. J. Ed. Psychol., 19, 381–388.

Hollingworth, H. L. '27. Mental growth and decline. New York: D. Appleton and Co. Pp. 396.

Holmes, S. J. '21. The trend of the race. New York: Harcourt, Brace and Co. Pp. 396.

Jennings, H. S. '24. Heredity and environment. Sci. Mo. 19, 225–238.

Kellogg, V. L. '23. Mind and heredity. Louis Clark Vanuxem Foundation, Princeton Univ. Press, Pp. 108.

Kent, G. H. '23. A combination mental test for clinical use. J. Appl. Psychol. 7, 246–257.

Lowie, R. H. '29. Are we civilized? New York: Harcourt Brace and Co. Pp. 306.

Ogden, R. M. '26. Psychology and Education. New York: Harcourt Brace and Co. Pp. 364.

Pearson, K. '10. Nature and nurture. The problem of the future. Eugen. Lab. Lect. Series, 6.

Perrin, F. A. C. and Klein, D. B. '26 Psychology: its methods and principles. New York: H. Holt and Co. Pp. 387.

Pintner, R. '29. Educational psychology. New York: Henry Holt and Co. Pp. 378.

Popenoe, P. '15. Nature or nurture? J. Hered., 6, 227–240.

Terman, L. M. '28. The influence of nature and nurture upon intelligence scores: an evaluation of the evidence in Part I of the 27th Yearbook. J. Ed. Psychol., 19, 362–373.

Thorndike, E. L. '13. Educational psychology. Vol. I: The original nature of man. New York: Teach. Coll. Pp. 277.

Whipple, G. M. '28. Editorial impression of the contribution to knowledge of the 27th Yearbook. J. Ed. Psychol. 19, 389–396.

Part Two, Chapter I

THE ANALYSIS OF ENVIRONMENT

The existence of environmental differences as a fundamental source of difficulty for mental testers has already been mentioned. When one considers that what a mental test actually proves is whether or not an individual has taken a share in his surroundings, the importance of knowing about those surroundings is evident. This is particularly true when groups with widely differing backgrounds are compared. When the psychologist tries to test primitive peoples for instance, the question of culture variations is paramount, and he must ask the anthropologist for appropriate material with which to devise tests. Most users of tests in this country still make Binet's assumption that experience in the activities tested is, on the whole, equal.[15] Almost everyone admits the desirability of tests which are independent of the amount of schooling or of the cultural background which a person has had, but so far it has been impossible to construct such a test because the *modus operandi* of many environmental factors is still unknown to us, and therefore we can neither account for them, nor offset them in a test situation. The result is that the most useful tests are those which most nearly approach the life situations for which they are to predict ability. In short, until the analysis of environment is seriously undertaken, mental tests will be more or less useless as tools of comparison, and their chief function will continue to be, as it is now, to classify men within homogeneous groups. (See Symonds' 24, and Chapter Two on "inter" vs. "intra" group testing, pp. 47-51.)

Studies dealing with the influence of environment on innate ability are usually more concerned with the relative importance of the two factors than with the influence of one kind of environment on a specific capacity. As is well known, such studies furnish conflicting evidence. Some writers believe that environment is more or less responsible for achievement; others regard intelligence as a causative factor in the pursuit of an occupation, and hence in the various cultural and social distinctions that surround different jobs.[16] The army tests revealed enormous differences in intelligence between the vocational groups, and this stimulated many workers to examine the relation existing between ability and work. The practical question is whether the occupations which are considered lowly according to academic standards, are causes or effects of inferior ability. Though most work on this subject has led to the conclusion that humble positions are held by the less intelligent, a recent study by Witty and Lehman ('30) finds that, statistically speaking, the majority of gifted children in America come from families with small incomes. Eventually it is hoped that case studies will be made to prove whether selection operates when people move from one occupational area to another. For instance, how do we know if it is the more intelligent of the Negroes who travel cityward—whether "blood seeks environment"? Ideally, all migrants should be tested before they move, and their children tested after them.[17]

The schools of a nation furnish a situation in which the two most important aspects of western environment—the family and education—can be analyzed. The first is studied more or less indirectly by methods which will be touched upon below. Because of the wide extent and standardized nature of our educational system, it could almost be called the "controlled factor" in studies of family influence. Walter Lippman ('22) points out that even after children have been exposed for six years to the American School —"the Great Leveller"— a significant relation exists between intelligence and socio-economic status. Kelley ('26) measured groups of eight, eleven and fourteen year old children with the Stanford Achievement Test and found that "nurture tends to eliminate idiosyncrasy". As to the second important aspect of environment, it is obvious that formal education provides the most incontrovertible proof of individual differences. We have seen that mental tests were originally scaled according to a pragmatic conception of ability as judged by the "common sense of the world". Scholastic marks and teachers' ratings are the most widely used criteria of intelligence for tests aiming to be independent of "training". There is no reason, however, why the equality of training provided by a school should not be used to measure native mental differences directly. Franzen ('22) presents data in which intelligence is isolated by a spelling test ("the influence of environment . . . . is of no consequence, since the subjects all had the same schooling, and home influence does not as a rule teach children to spell"), and Kelley ('27) points out that ninety per cent of a good intelligence test measures the same thing as an achievement test. Adams ('23), while admitting that if tests had universal validity and "could be applied with the same effects on all sorts and conditions of men", they would be "more convenient" than they are now, says that "after all, school attainments are as much a part of real life as are attainments acquired outside of school, and intelligence may as well be tested by one as by the other". Many studies have been made to determine the influence of schooling and special training on mental tests,[18] leading to the general conclusion that little connection exists between gain in I.Q. and length of school attendance, while only temporary improvement follows upon special coaching. The controversy which has centered around the "constancy of the I.Q." should be mentioned[19]; to discuss it fully, however, together with the many other studies of different environmental influences,[20] would lead too far afield.

Schooling

Efforts to deal with the factor known as "socio-economic status" (which, with schooling, comprises most of the cultural influences to which our children are subjected) began almost as soon as the testing movement was started. Early studies were made by Decroly and Degand ('10), Morlé ('11), Hoffman ('14), Yerkes and Anderson ('15) and Bridges and Coler ('17). Thoughtful workers sought to establish the principle that children coming from "excellent" neighborhoods should not be judged by the same norms as those from "poor" homes.[21] In 1913 Meumann suggested an age scale giving, for each year, tests of intelligence, development, and environment. (Such an idea suggests fruitful research, but nothing has apparently been done with it). A good summary of later studies dealing with home environment is given by Burdick ('28). H. E. Jones, referring to

Home Life

the importance of obtaining a picture of the home when studying children's social adjustments says:

> "At the present time it is difficult to reduce any significant phases of this picture to quantitative terms, because of the crude state of our methods of social measurement". (Conference of the Child Development Committee of the National Research Council, Toronto, 1929.)

The casual narrative description of the home which the individual worker used to make, has given way to more standardized methods of report, such as inventories of home equipment, or rating scales based on occupation, income and "culture." [See, besides the well known Barr and Taussig scales, VanDenberg ('11), Kornhauser ('18) Chapin ('20), Counts ('22), Williams ('20) and Mackaye ('29)]. Questionnaires for children to fill out have been made up by Holley ('16) and Sims ('28). Most of these workers have developed some sort of "home index", the criteria for which range all the way from the number of books in the library to the opinion of the janitor of the apartment house in which the family lives! Burdick ('28) has evolved a group test, the questions of which are designed "to stimulate answers significant of the environment in which a child lives" in such a way that he will not suspect the purpose of the examiner. Thus, the child must mark "true" or "false" such phrases as: "One shall assist the hostess by stacking the dishes"; and instead of "Do you have a newspaper in your home?" he is asked "What is your favorite newspaper?" The employment of such techniques goes a long way towards controlling the personal equation, yielding scores that can be used statistically, with a consequent saving in time and money. With further refinement and extensive application of these methods, it may eventually be possible to describe in detail any desired sample of social background.

Many retest studies have been made to discover the effect of improved surroundings on the same people (see, for example, Teagarden '27, Barrett and Koch '30, and the articles by Rogers, Hildreth and Goodenough in the 27th Yearbook of the National Society for the Study of Education).

Retests (see also page 32)

Freeman ('28) views the results of some of these researches as indications that the I.Q. may be raised as much as ten points by an ameliorated environment, although he admits that all findings are not consistent. For instance, Rogers and her Bryn Mawr colleagues found that when children who had been living in an admittedly "poor" milieu were placed in a well managed institution for over a year, they did far better in their school achievement, but showed no signs of improvement on the Binet. On the other hand, Hetzer and Wolf ('28) seem to prove that good care during the first year of life does improve a baby's intelligence as judged by the "Baby Tests". These same tests were used by Myrtle McGraw on Negro and white infants, ranging in age from two to eleven months. She found the usual white superiority, and although these results may be more important than those from racial comparisons with older children, she herself admits that she could not pair the babies against each other because of "social difficulties"— i. e. because certain environmental factors could not be isolated. (Reported at the Annual Meeting, New York Chapter of the American Psychological Association, 1930).

Conclusion

The need for analyzing specific elements of an environment (as distinguished from general "superiority" or "inferiority") is widefelt. Antipoff ('28) shows that "social variation can mask all other kinds of differences and lead to grave error in the establishment of norms" (see also page 41 of this volume). Since the home is the most important environment, at least for the young child, the relationship between children's characters or abilities and such factors as vocabulary, toys, travel, playmates, books, radios, urban versus rural life, and parents' intelligence and vocation should be carefully worked out (see Schwesinger '26 and Van Alstyne '29). Walter Lippman ('22) says: "The whole more or less blind drama of childhood, where the habits of intelligence are formed, is concealed in the mental test."; while according to Kimball Young ('24), the problem of heredity and acquired characteristics must be solved by a careful study of "the subtle influence of the early conditioned responses which have been built up in the home, the neighborhood, the playground and the schoolroom."

But even with the most complete knowledge possible, there will always remain "intangible elements and combinations of influences which give a home its quality",[22] and it is doubtful whether these can ever be expressed in quantitative terms. Let us take, for instance, two children with the same schooling and social background in a western rural community, who are given an intelligence test. One might have a father of the pioneer type, possessing keen practical ability and knowledge of men, yet with a peculiar turn of mind which leads him to scorn "book learning". The parents of the other child, either because of chance experience or inclination, might recognize the benefits of education, and in countless small ways encourage him to learn all he can. The probability is that the second child will score higher than the first, but who would dare interpret this as meaning that he is more "intelligent"? It is for this reason that tests can never yield complete measurements: as Johnson ('29) points out, the property to be tested depends only *partially* on the property employed to test it. In our quest for measurements of the inherited bases of mental aptitudes, we are therefore thrown back on the necessity of using tests to compare groups of individuals within specific ranges. It is only after this has been done many times that the analysis of different environments can progress, with the ultimate result that tests will not be limited to the definition of mere statistical "tendencies", but will be really useful instruments for individual analysis.

BIBLIOGRAPHY

References cited in Part II, Chapter I.

The Measurement of Environment

(See Burks '28 pp. 268-282 and 296-326 for annotated supplements to this Bibiography.)

Adams, J. '23. Educational implications and the I.Q. Australasian J. of Psychol. and Phil. 1, 177–190.

Aldrich, S. C. '27. The intelligence of high-school pupils. Sch. Rev., 35, 699–706.

Antipoff, H. '28. L'évolution et la variabilité des fonctions motrices. Arch. de Psychol., 21, 1–54.

Arthur, G. '26. The relation of I.Q. to position in family. J. Educ. Psychol., 17, 541–550.

Bagley, W. C. '23. Do good schools pay? J. Nat. Ed. Assoc., June 11, 221–216.
—'25. Determinism in education. Baltimore: Warwick and York, Pp. 192.
Barrett, H. E. and Koch H. L. '30. The effect of nursery school training upon the mental test performance of a group of orphanage children. J. Genet Psychol. 37, 102–122.
Blanchard, P. and Paynter, R. H. '27. Socio-psychological status of children from marginal families. The Family, 8, 3–10.
Book, W. F. '22. The intelligence of high school seniors. N. Y. The MacMillan Co. Pp. 371.
Bradford, E. J. G. '25. Can present scholastic standards be maintained? Forum Educ., 3, 186–198.
Bridges, J. W. and Coler, L. E. '17. The relation of intelligence to social status. Psychol. Rev., 24, 1–31.
Broom, E. '27. Constancy of the I.Q. Sch. and Soc., 25, 295–296.
Burdick, E. M. '28. A group test of home environment. Arch of Pspchol., No. 101. Pp. 115.
Burks, B. S. '28. A summary of the literature on the determiners of the intelligence quotient and the educational quotient. 27th Yrbk. Nat'l. Soc. Stud. Educ., Pt. II, 248–353.
Burt, C. '21. Mental and scholastic tests. London: County Council, Pp. 432.
Cattell, P. and Gaudet, F. T. '30. The inconstancy of the I.Q. as measured by repeated group tests. J. Educ. Res., 21, 21–28.
Chapin, F. S. '20. Field work and social research. New York: Century Co., Pp. 224.
Chapman, A. E. '24. The effect of school training and special coaching on intelligence tests. Forum Educ., 2, 172–183.
Chapman, J. C. and Wiggins, D. M. '25. Relation of family size to intelligence of offspring and socio-economic status of family. Ped. Sem., 32, 414–421.
Chauncey, M. R. '29. The relation of the home factor to achievement and intelligence test scores. J. Educ. Res., 20, 88–90.
Chen, H. S. '28. The comparative coachability of certain types of intelligence tests. Columbia University: T. C. Contrib. to Educ., 338, Pp. 101.
Colvin, S. S. and MacPhail, A. H. '24. Intelligence of the seniors in the high schools of Massachusetts. Bur. Educ. Bull. No. 9. Pp. 66.
Conklin, A. M. '30. Families of intellectually gifted students. The Family 11, 99–106.
Cornell, E. L. '27. Taking the dogma out of the I.Q. Mental Hygiene, 11, 804–810.
Counts, G. S. '22. Selective character of American secondary education. Supplementary Monograph, Univ. of Chicago, Pp. 162.
Courtis, S. A. '26. The influence of certain social factors upon scores in the Stanford achievement tests. J. Ed. Res., (a) 13, 311–323 and (b) 14, 33–42.
Davis, R. A. '28. Some relations between amount of school training and intelligence among Negroes. J. Educ. Psychol., 19, 127–130.
Dearborn, W. T. '24. Repeated measurements of the physical and mental development of school children. Sch. and Soc. 20, 515–518.
Decroly, O. '26. At what age does intelligence cease to develop? And intelligence in relation to different social classes. Mental Welfare, 7, 9–16.
— and Degand, J. '10. La mesure de l'intelligence chez les enfants normaux d'aprés les tests de Binet et Simon. Arch de Psychol., 9, 81–108.
Dvorak, A. '24. The relation of I.Q. to the prognosis of special class pupils. Sch. and Soc., 19, 736–744.
Foster, J. C. '19. A case of intellectual development despite enforced seclusion. J. Appl. Psychol., 3, 167–171.
Franzen, R. '22. The accomplishment ratio. T. C. Col. Univ. Contrib. to Educ., No. 125. Pp. 59.
Freeman, F. N. '28. An evaluation of the evidence in Part I of the Yearbook, and its bearing on the interpretation of intelligence tests. J. Educ. Psychol. 19, 374–380.
Fryer, D. '22. Occupational intelligence standards. Sch. and Soc., 273–277.
Fukuda, I. '25. A survey of the intelligence and environment of school children. Am. J. Psychol., 36, 124–139.
Furfey, P. H. '28. The relation between socio-economic status and intelligence of young infants as measured by the Linfert-Hierholzer scale. Ped. Sem., 35, 478–480.

Gates, A. I. '28. Observed facts and theoretical concepts. J. Ed. Psychol., 19, 381–388.

— and LaSalle, J. '23. The relative predictive value of certain intelligence and educational tests, together with the effect of educational achievement upon intelligence scores. J. Educ. Psychol., 14, 517–539.

Gesell, A. and Lord, E. E. '27. A psychological comparison of nursery school children from homes of low and high economic status. Ped. Sem., 34, 339–356.

Goodenough, F. L. '26. Racial differences in the intelligence of school children. J. Exper. Psychol., 9, 388–397.

—'28a. A preliminary report on the effect of nursery school training upon the intelligence test scores of young children. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I, 361–369.

—'28b. The relation of the intelligence of pre-school children to the occupation of their fathers. Amer. J. Psychol., 40, 284–294.

— and Shapiro, G. '28. The performance of pre-school children of different social groups on the Kuhlman Binet tests. J. Educ. Res., 18, 356–362.

Gray, P. L. and Marsden, R. E. '26. The constancy of the intelligence quotient. Final results. Brit. J. Psychol., 17, 20–26.

Hetzer, H. and Reindorf, B. '28. Spranchentwicklung und Soziales Milieu. Zsch. f. Angew. Psychol., 29, 448–462.

— and Wolf, K. '28. Baby tests. Zsch. f. Psychol. 107, 62–104.

Hildreth, G. '26. Stanford-Binet retests of 441 school children. Ped. Sem., 33, 365–386.

—'28. The effect of school environment upon Stanford-Binet tests of young children. 27th Yrbk. Nat'l Soc. Stud. Educ. Pt. I, 355–359.

Hirsch, N. D. M. '28. An experimental study of the East Kentucky Mountaineers. Genet. Psychol. Monog., 3, 245 minus 189.

Hoffman, A. '14. Vergleichende intelligenz Prüfungen an Vorschülern und Volkschülern. Zsch. f. Angew. Psychol., 8, 102–120.

Holley, C. E. '16. "Relationship between persistence in school and home conditions." Reprint from 15th Yrbk. Nat'l. Soc. Study of Educ. Chicago: 15, Pt. II, Pp. 115.

Hollingworth, L. S. '26. Gifted children; their nature and nurture. New York: Macmillan Company. Pp. 363.

Hurlock, E. B. '25. The effect of incentives upon the constancy of the I.Q. Ped. Sem., 32, 422–434.

Johnson, H. M. '29. Science and Sorcery in mental tests. Forum, 82, 366–372.

Jones, D. C. and Carr-Saunders, A. M. '26–'27. The relation between intelligence and social status among orphan children. Brit. J. Psychol., 17, 343–364.

Kelley, T. L. '26. The influence of nurture upon native differences. New York: The Macmillan Company. Pp. 49.

—'27. The interpretation of educational measurements. Yonkers: World Book Company. Pp. 363.

Kempf, G. A. and Collins, S. D. '29. A study of the relation between mental and physical status of children in two counties in Illinois. U. S. Pub. Health Reports, 44, 1743–1784.

Kiefer, F. A. '29. Manual motor correlation in superior children. J. Appl. Psychol. 13, 357–371.

Kornhauser, A. W. '18. Economic standing of parents and intelligence of school children. J. Ed. Psychol., 9, 159–164.

Lennes, N. J. '27. Whither democracy? N. Y.: Harper. Pp. 370.

Lentz, T. '27. Relation of I.Q. to size of family. J. Educ. Psychol. 18, 486–496.

Lippman, W. '22. Tests of hereditary intelligence. New Rep. 32, 328–330.

Macdonald, H. '25. The social distribution of intelligence in the Isle of Wight. Brit. J. Psychol., 16, 123–129.

MacKaye, D. L. '29. Interrelations of speech and intelligence. Amer. J. Sociol. 35, 353–368.

Matthew, J. A. and Luckey, B. M. '28. Notes on factors that may alter the intelligence quotient in successive examinations. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I, 411–419.

Meumann, E. '13. Vorlesungen. Leipzig: Englemann.

Morlé, M. '11. L'Influence de l'état social sur le degrè de l'intelligence des enfants. Bull. Soc. Libre Educ. Psychol. de l'enfant. 12, 8–15.

Nettles, C. H. '26. A study on the constancy of the I.Q. Educ. Res. Bull., Los Angeles schools, 6, 9–10.

Pintner, R. and Paterson, D. G. '17. A scale of performance tests. New York: Appleton, Pp. 218.

Pressey, S. L. and Thomas, J. B. '19. A study of country children in (1) a good, (2) a poor farming district by means of a group scale of intelligence. J. Appl. Psychol. 3, 283–286.

Pressey, L. W. '20. The influence of (a) inadequate schooling and (b) poor environment upon results with tests of intelligence. J. Appl. Psychol., 4, 91–96.

Probst, C. A. Extent and range of information possessed by kindergarten children in Minneapolis schools. (Unpublished Master's thesis at Univ. of Minnesota.)

Randall, F. B. A. '27. A study on the constancy of the I.Q. Sch. and Soc., 26, 311–312.

Rugg, H. O. and Colloton, C. '21. Constancy of the Stanford Binet I.Q. as shown by reteests. J. Educ. Psychol. 12, 315–322.

Schwesinger, G. C. '26. The social-ethical significance of vocabulary. T. C. Contrib. to Educ. Col. Univ. Pp. 73.

Sims, V. M. '28. The measurement of socio-economic status. Bloomington, Ill.: Publ. Sch. Publ. Co. Pp. 33.

Sirkin, M. '29. The relation between intelligence, age and home environment of elementary school pupils. Sch. and Soc. 30, 304–306.

Slocombe, C. S. '27. Why the I.Q. is not and cannot be constant. J. Educ. Psychol. 18, 421–423.

Stern, W. '24. The psychology of early childhood. Eng. Trans. 3rd. Ed, New York. Henry Holt and Company. Pp. 557.

Stoke, S. M. and Lehman, H. S. '30. Intelligence test scores of social and occupational groups. Sch. and Soc. 31, 372–377.

Strong, A. C. '13. 350 white and colored children measured by the Binet-Simon scale. Ped. Sem. 20, 485–513.

Stroud, J. B. '28. A study of the relation of intelligence test scores of public school children to the economic status of their parents. J. Genet. Psychol. 35, 105–110.

Sutherland, H. E. G. and Thomson, G. H. '26. The correlation between intelligence and size of family. Brit. J. Psychol. 17, 81–92.

Symonds, P. M. '24. Verbal versus nonverbal tests as valid intelligence tests in Hawaiian Schools. Sch. and Soc. 20, 248–249.

Teagarden, F. M. '22. "The constancy of the I.Q." again. J. Educ. Psychol. 13, 366–372.

—'27. Change of environment and I.Q. J. Appl. Psychol. 11, 289–296.

Terman, L. M. '16. The measurement of intelligence; an explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale. Boston: Houghton Mifflin, Pp. 362.

Van Alstyne, D. '29. The environment of three year old children; factors related to intelligence and vocabulary tests. T. C. Contrib. to Educ. Col. Univ., 339, Pp. 108.

VanDael, J. '29. De invloed can het sociale milieu op de ontwikkeling dar intelligentie. (The influence of social environment on the development of intelligence.) Mensch en maat schappij, 5, 339–341.

Van Denberg, J. K. '11. Elimination of students in public secondary schools. Col. Univ. T. C. Series. No. 47. Pp. 206.

Wechsler, D. '26. On the influence of education on intelligence as measured by the Binet-Simon tests. J. Educ. Psychol., 17, 248–257.

Weintrob, J. and Weintrob, R. '12 The influence of environment on mental ability as shown by Binet-Simon tests. J. Ed. Psychol. 3, 577–583.

Wells, G. R. '23. The application of the Binet-Simon tests to white and colored children. Psychol. Monog., 32, 52–58.

Willard, D. W. '22. Native and acquired mental ability as measured by the Terman group test of mental ability. Sch. and Soc., 16, 750–56.

Williams, J. H. '20. Whittier scale for grading home conditions. Bull. No. 7, Whittier State School, California. Pp. 21 plus score sheet.

Wooley, H. T., '25. The validity of standards of mental measurement in young childhood. Sch. and Soc. 21, 476–482.

Yerkes, R. M. '21. Psychological examining in the United States Army. Mem. Nat. Acad. Sci., 15. Pp. 490.

— and Anderson, H. '15. The importance of social status as indicated by the results of the point scale method of measuring mental capacity. J. Educ. Psychol. 6, 137–150.

Young, K. '24. The history of mental testing. Ped. Sem. 31, 1–48.

Part Two, Chapter II

STUDIES OF MENTAL INHERITANCE

Introduction

While recognizing that "heredity and environment, as integral parts of the developing organism" cannot be separated, it is legitimate to try to discover "whether or not mental traits are transmitted by Mendelian factors, and if so, through interminable research, slowly to identify those factors, locate them in the chromosomes, and determine their degrees of dominance."[23] Before this can be done, however, there must be developed proper tools of mental measurement. This is not merely a matter of constructing a tool, as in anthropometry, because the *use* to which a mental test is put is as important as the test itself. Kelley ('30) points out that the psychologist is interested in the growth phenomena which take place between germ cell and germ cell, while the biologist deals only with genetic structure and its transmission from parent to offspring. The problem of measuring innate mental traits is enormously complicated by the fact that these growth phenomena are not limited to "maturation processes", but necessarily include the entire learning process. The psychological traits of man are greatly altered by circumstances, and hence cannot be measured directly as can be height or eye color, and even various mental abnormalities.[24] But if circumstances can be adequately controlled, there is no reason why psychometry should differ qualitatively from anthropometry;[25] indirect measurement merely means that two factors, one of them controlled, must be considered instead of a single factor.

The use of the mental test method in social research dealing with mental heredity can be said, for convenience' sake, to have taken the following three courses, each of which will be briefly described in this chapter: (1) Time studies or "retests"; (2) Family resemblance studies; and (3) Group comparisons. It should be noted that these three procedures are neither strictly comparable nor sharply distinguishable. The first is really a *technique* which we have already seen applied in studies of environment. It can merely be touched upon below, for the literature on retests is too vast to be adequately dealt with in this volume. The second is a definite "field" of research, dealing directly with the problem of heredity, and an attempt will be made to summarize the most important studies on the mental likeness of relatives. The third should hardly be separated into a class by itself, since *all* studies in which mental tests are used (even the individual clinical diagnoses) depend ultimately on the comparison of groups. Of the many categories which serve to differentiate human beings into groups, race has been chosen for discussion below because inter-racial testing illustrates nearly all the aspects, favorable and unfavorable, of inter-group testing in general. This explains why a seemingly disproportionate amount of space is devoted to racial studies.

RETESTS

The term "retest", "time" or "follow up method" has already been mentioned (page 25). It means testing the same individual or group at different times, when the environment has been varied or held constant. It can be used to test innate ability in two ways: first, directly, by retesting at long intervals, and second, inferentially, by measuring learning ability by frequent retests. The former method is admittedly the soundest, but is still purely theoretical. If, as Wallis ('26) asserts, the permanence of a trait attests its biological origin, then families must be measured throughout many generations, regardless of changing circumstances, to discover which characteristics are lasting. (See also Appendix I, No. 26) Woodworth ('10) and Garth ('21a) have outlined ideal experimental set-ups whereby two equal groups would be reared amid controlled surroundings. Were such schemes humane and possible, we might, according to Woodworth, be able to tell whether these "two groups alike in numbers, heredity, and environment, would remain alike, or progress at equal rates." Needless to say, such schemes are practically out of the question; indeed, the labor and expense involved in following up and retesting school groups even at short intervals renders the "time method" one of the most wearisome techniques in the study of mental heredity.

Frequent retests (i.e. measurement of the learning capacity) is a recognized technique in many fields. Burks and Kelley ('28) say:[26]

> "We need more studies . . . . such as the measurement of racial differences with the possible effects of environment eliminated through experimental control, and studies employing rigorously matched control groups, to measure the week-by-week improvement on mental and achievement tests, of foreign-speaking subjects as they learn English." p. 29.

Mental testers all over the country are retesting their subjects because they recognize the difference between initial ability and improvability. One of the fundamental problems of genetic psychology is to distinguish between maturation (general or specialized) and learning. Since there must be an unlearned foundation for every learned response, we can say that proficiency in a test depends partly on ability and partly on the "technique, information and adjustments to the task which may be acquired during practice." (Gates and Taylor '25). It is therefore of utmost importance that the tester of "innate ability" should know which performances show the "highest correlation between original and final status", because obviously they are the most sensitive to innate capacity.[27] (See also pages 45 and 46 of this volume)

FAMILY RESEMBLANCES

Expositions of family resemblance in the form of genealogies are widely known. Beginning with Galton's "Heredity Genius", where certain stocks which have produced men of genius are examined, impressive pedigrees of both famous and infamous lines have been appearing. (See, for instance, Winship '00, Ellis '04, Woods '06, Dugdale '10, Estabrook '16, Cattell '21 Brimhall '23, Pearson '24 and Kretschmer '30). Besides this, the relatives of gifted children have also been studied with reference to Who's Who and the Hall of Fame. (Woods '13, Visher '25 and Terman '25). The general verdict of these investigations, despite the briefs of DeCandolle ('73) and Constable ('05) for environmental influence, is that intelligent children *on the whole* tend to have intelligent ancestors and relatives (although it cannot be said that "like produces like").

Another way of studying family resemblance is to measure related persons by means of correlation. According to Pintner ('29), if statistics can prove that relatives are as similar mentally as they are physically (the latter resemblance being uninfluenced by similar environment), "heredity" must be a causal factor in mental as well as physical traits. He gives the following general median correlations: .5 between siblings for such traits as eye and hair color, and cephalic index; .78 for the intelligence of pairs of twins; .51 for that of siblings; and .31 for that of cousins. He concludes that these figures "fit in with the assumption that intelligence is inherited in exactly the same way as physical characteristics". Willoughby ('28) obtained correlations of .43 and .39 for fraternal and paternal resemblances respectively, and the coefficients of H. E. Jones ('28) for the same comparisons were .49 and .5. Banker computed a Students' Ability Index for teachers' marks, and with it found approximately the same correlations as those for physical characters. He infers that this "seems clear evidence that a complex of fundamental factors having genealogical relationships" is being dealt with.

Granting that relatives are as much alike intellectually as they are physically, the question of environmental influence still remains unsettled. Even though the similar family background may not affect such traits as eye color and even the shape of the head (though this last might be disputed), it very probably operates to make two siblings more alike in their ideas and interests than two random brothers or sisters. Thorndike ('28) finds a correlation of .60 for the intelligence of unselected brothers and sisters, and Whipple ('28), comparing this with Pearson's .52 for the physical resemblance of siblings, says:

> "Assuming that intelligence and physical traits exhibit about identical hereditary tendencies.... it would follow that the similarity of the environment that exists between siblings two to four years apart in age in an American family today is sufficient to raise the natural correlation of their intelligence scores from .52 to .60...."

A few studies have sought to show that siblings are just as unlike in mental traits which presumably are affected by training as in those which are not (see, for instance, Starch '19, p. 82). Thorndike ('05) found that twins fail to grow more alike with the continuance of a common environment; that they are more similar than ordinary siblings, and that they show no greater resemblance in subject matter susceptible to training. In the case of parent-child correlations, Pearson ('10) deduced that the coefficient for nature against nurture was as .51 is to .03. He declares: "I think it is quite safe to say that the influence of environment is not one-fifth that of heredity, and quite possibly not one-tenth."

The above studies, interesting though they may be, do not eliminate the influence of common family life, and hence cannot furnish crucial evidence on the heredity-environment question. Far more convincing as techniques for *isolating* the effects of each of these factors are studies of so-called "identical" twins and of foster children. Morgan ('29) calls the former "the most available materials for a scientific examination of the role of nature versus nurture in human psychology". Many studies of identical twins have been made, one of the earliest of which was Galton's history

Identical Twins

of twins in his "Inquiries into Human Faculty" ('83). After finding that changed surroundings had little effect on the mental and physical similarities of several pairs of twins, he concludes:

> "There is no escape from the conclusion that nature prevails enormously over nurture when the differences of nurture do not exceed what is commonly to be found among persons of the same rank of society and in the same country."— p. 241.

Identical twins who have been reared apart were studied by Muller ('25) and Newman ('29). The former found that diversity of environment had little effect on intelligence tests scores but produced real differences in temperament. Newman on the other hand found quite the reverse: the emotional reactions of each pair of twins were similar, while their intelligence test scores varied widely. Jennings ('30) comments thus on these two studies:

> "The results . . . . agree thoroughly with the conclusion . . . . that both the genetic constitution and the environment deeply influence mental and temperamental characteristics; and that effects produced in one case by genetic constitution may be produced in another case by environment".

Merriman ('24) has tested twins and obtained correlations on the Binet of .80 for like sex, and .50 for unlike sex pairs. Tallman ('28) measured the difference between pairs of (1) identical twins (2) ordinary twins and (3) siblings. She found, as was to be expected, that the difference was small in case (1), somewhat greater in (2), and very large in case (3).[28] Gesell and Thompson ('29) worked with a pair of identical infant twins in order to watch the effect of training and no training on the "same" individual. Their conclusion was that "training does not transcend maturation". A pair of Siamese twins was measured by Koch ('28), and the divergence between them was found to be as large as that usually found between ordinary identical twins. Whipple ('28), commenting on these results, points out that identity of environment can hardly be more closely approached than by physically hooking two children together. Finally, Haldane ('29) cites Lange's *Verbrechen als Schicksal* ("Crime as Destiny") as "conceivably the most important book of this century", because it shows how differences of environment and free will together saved only three out of thirteen "identical" twins from imitating their criminal brother and sister.

Foster Children

The study of foster children makes use of a situation in which, according to Burks, the influence of nurture alone can be compared with the combined effect of nature and nurture. Terman ('28) cites the Stanford and Chicago foster-children studies in the Twenty-seventh Yearbook of the National Society for the Study of Education (Burks '28 and Freeman *et al* '28) as "the only ones (in the Yearbook) that have made a serious attempt to disentangle heredity and environment". The Stanford results showed a correlation of .56 between parents and their own offspring, and, in the case of children adopted when only a few months old, coefficients of .09 and .23 for foster fathers and mothers respectively. Statistical analysis of these data, coupled with the fact that selective placement was as far as possible eliminated, leads the author to consider that "if the measureable home environment of foster families were made constant, the variance in I.Q.'s of foster children would be reduced by only seventeen per cent." A different statistical treatment of these results made it appear that a child's I.Q. is raised or depressed six points by a home environment one sigma above or below the average. The Chicago study, on the other hand, seems to show that environmental influence is far more important than Burks thought, since it was found that siblings who were separated and reared in different homes resembled each other less than siblings in general. A "substantial relationship" was found between the character of the home and the intelligence of the children, so that Freeman concludes that the influence of environment alone gives a correlation of about .45 which is raised to .60 if the influence of parental heredity is added. In any case, both these studies show that "in the best homes, more things are learned which influence Binet test scores than are learned in the poorest." (Gates '28).

However, granting that extreme differences in environment may cause extreme differences in the organism,— a "normal organism or a monstrosity, ... a living organism or none at all" (Holmes '21) — we are bound to admit with Burks ('29) that good and bad homes may raise or lower the I.Q. only about twenty points, "while heredity can produce alike the idiot of twenty and the genius of 200 I.Q.". To quote Holmes again:

> "The ordinary differences of environment met with in the life of people of much the same mental status apparently fail to produce changes in the personality of human beings as great as commonly met with in the children of the same parents."—'21, p. 25. (See also pages 21, 34, 252 and 254 of this volume.)

GROUP COMPARISONS

The third line of attack in the study of mental heredity will be dealt with as we have said by using inter-racial testing as a concrete example of group comparisons in general. It is assumed that the problems attending race comparisons are so numerous and clear cut that they provide good illustrations of the problems attending other kinds of group comparisons in which either a hereditary or environmental factor is supposedly isolated — such as studies of sex or urban-rural differences.

Introduction to Discussion of Race Comparisons

The existence of different races on this earth creates a situation where nature exhibits the workings of heredity, and consequently racial crossing has become one of the geneticist's best tools.[29] Before the laws of segregation and recombination can be worked out, there must exist a means of measuring innate mental traits. To repeat yet again, the best way to do this at present is to compare groups of individuals so that differences of race or environment can be controlled. Once it were established that a certain test measured intelligence, say, despite language differences, that test could be used to measure parents and offspring in racial crosses.

Since the validity of all group comparisons, and especially of inter-racial testing, depends on the use of equivalent measuring units, it stands to reason that tests must be equally fair—i.e. equally novel or familiar—to the groups in question. (See page 42). Thorndike ('26) has pointed out that this can be accomplished either by testing infants before environmental differences have had their influence, or else by equating the measuring instrument in some way. Since this study is not concerned with the former method, the rest of this chapter will deal with the problems involved in making valid comparisons between children or adults of different races. Two testing principles will be discussed: "universal measurement" and so-called "intra-group scaling".

Racial differences in such things as climatic adjustment, resistance to disease and material culture are observable facts, and anthropometry has succeeded fairly well in measuring physical types that could be classed as racial. If we assume that race (defined by Marett as "inherited breed" and by Brigham[11] as "suspected similar ancestry") depends on more than physical characteristics—that it relates to the mind as well as to the body—it is legitimate to try and measure racial mentality. Wissler ('23) says: "Races are differentiated by descent, hence it follows that their innate equipments will differ". That they do differ, and that the differences take the form of inferiorities when other races are compared with ours, is an assumption around which controversy has raged for the last few decades.[30] (Even Galton took for granted, without objective tests, that Negroes were two grades below whites on his scale of fourteen grades. For a discussion of "real" racial differences, see Appendix I, No. 80.)

Comparative studies of racial differences can be divided into the testing of races living within our own civilization, and the testing of primitive peoples in their native surroundings. The practical importance of analyzing the racial elements composing our citizenry, besides a theoretical interest in the relation between race and nationality, has led many investigators to make studies of the first kind. The testing of primitive peoples, on the other hand, has been undertaken for different reasons, the origin of which could, in many cases, be traced to the civilized person's general curiosity toward any "savages" which adventurers bring into his ken. More specifically, the anthropologist desires to secure information about the indigeneous life of fast vanishing tribes; the geneticist is concerned about the outcome of race mixtures; and the sociologist is interested in the effect of different environments on the same racial stock—as, for instance, when Negroes were transported from Africa to America.

DIFFICULTIES IN INTER-RACIAL TESTING

In using intelligence tests as a means of comparing races, certain major difficulties immediately appear, whether the testing is carried on in New York City or in the bush. The primary problem is that of defining "race" as a biological concept, and it is such a complex question that even anthropologists are not agreed upon it. Blackwood ('27) says:

Definition of "race"

> "At present, even if we had at our disposal mental tests guaranteed to measure intelligence accurately, we could not claim that they would, with equal accuracy, measure *racial* intelligence. We are, in point of fact, dealing with two unknown quantities, race and intelligence, the interrelations of which, expressed in terms of heredity, are also unknown."— p. 6.

Hankins ('26) summarizes the views of various writers on race, (p. 263) and concludes that the term implies:

> ". . . . an ensemble of physical traits, each more or less variable, all inherited as an ensemble within their limits of variation, and sufficiently distinctive when taken together to mark off their possessors from other members of the human species."

A clear definition of racial divisions according to physical criteria is given by Hooton ('26), who points out that the first step in discovering whether races differ psychologically is to segregate pure racial types. (See also Estabrooks '28b). This cannot be done, to quote Blackwood again, until "many series of measurements and observations have been made—and light has been thrown upon obscure problems of human heredity." Thus we have practically a closed circle.

Sampling

Another problem is that of obtaining representative samplings. The two foundations of anthropological comparisons have been given as, first, the average of the racial group, and second, the high and low extremes. The distinction between these two methods of comparison must be kept clear, otherwise there is danger of "selecting downwards". This is best expressed by Brigham ('23):

> "To select individuals who have fallen behind in the struggle to adjust themselves to the civilization which their race has built . . . . to select children of equal education, age for age, in two groups (say whites and Negroes) is to sample either superior Negroes or inferior whites." (p. 194. See also Peterson and Lanier '29, Concl., and page 71 of this volume).

This necessity for obtaining comparable averages has also been stressed by Drs. Davenport[11] and Porteus and Babcock ('26). It does not obscure the fact, however, that in many cases the upper limit of a group is as important as the average. By computing both the range and the mean of two races, the per cent of the lower group which reaches or exceeds the average of the higher can be calculated, and most authorities are willing to attribute whatever difference is thus establised to race.[31] Says Hankins ('26):

> "To think of a race as having certain well defined traits without at the same time taking account of its variability is to leave out an essential datum for accurate thinking. (p. 261). The essential nature of race differences is one of differences in the mathematical proportions of the individual variations of which each is composed." (p. 293).

Temperament and Speed

A third difficulty is that the personality traits of the several races vary widely. The stolid Bohemian, the slow and accurate Indian, the impatient Italian, the nervously alert Hebrew and the easygoing Negro have all been measured with instruments that take little or no account of such temperamental differences.[32] It is by now an established fact that disposition affects test performance; Thorndike ('26 Chap. 16) shows how "original intellect" is conditioned by "original interest", in other words, temperamental "bent" or motivation. (See also page 16 of this volume). So-called racial differences in temperament may be biologically determined, and then again they may depend to a large extent on culture. In either case, they introduce a factor which, unless it is controlled, may well invalidate many comparisons of racial intelligence. Speed alone has often been considered as a temperamental trait. Gaw ('25) says:

> "Since so many performance tests are scored on speed, it follows that the subject's method of work—resulting from temperamental factors—is more dependent in these tests than in the Binet Scale on ability to work under pressure, and to withstand distraction by inessentials in the material, or by noises in the room."

Klineberg ('28) after studying speed variations among Indians, Negroes and whites, concludes that in performance tests, whites are superior to Indians and

Negroes, largely, if not entirely, in score for *time,* and that this greater speed is probably due more to environmental than to racial causes. The study made by Peterson, Lanier and Walker ('25) shows that white subjects are quicker than colored ones in certain tests, and stresses the importance of mental "set". (See also page 47 of this volume.) Mead[11] and Blackwood ('27), both anthropologists, point out that primitive subjects are usually not accustomed to a situation in which they must hurry. Blackwood hints at the possibility that our "perpetual driving by the clock", instead of being a sign of superior intelligence, might, in the eyes of the native, denote nervous instability. She concludes that if some intellectual performances can be measured largely in terms of quality or power, thus omitting speed, the anthropologist should use tests of that kind.

The relation of speed to intelligence is a question on which many of the best minds in psychology are occupied. Here, again, conclusions differ diametrically: On one side are those who find no connection between quickness and intelligence, such as Bernstein ('24) and Highsmith ('25). On the other side are those who assert that speed may often be taken as an indication of intelligence. Ruch and Koerth ('23) found a high correlation between speed and power in the Army Alpha test.[33] Peak and Boring ('26) make the statement "Speed of reaction is an important, and probably the most important factor in individual differences in the intelligent act". McFarland ('30) has shown that an individual's speed is a constant factor throughout a wide range of mental performance, while Kelley suggests that "mental speed" may be an "independent trait". (See page 15 of this volume.) Finally, Thorndike ('26) concludes that the correlation between speed and level of intelligence is so low that not much weight should be attached to the former except where it measures speed of learning.

Variations of culture between racial groups constitute perhaps the greatest difficulty for inter-racial testing. "Social inheritance" is such a powerful factor that, unless it is taken into account, observed differences between races can never be taken to mean innate differences. The culture of the Finnish district in Chicago might not contrast so strikingly with Middletown's as would that of the Solomon Islands, and yet its effect on test performance might be just as great.

Material Culture Differences

A case was reported to the writers of an immigrant woman from an East Central European country, who was given a pictorial test consisting mainly of drawings of household objects. She seemed intelligent enough in everyday life, and yet she could not recognize pictures of the pots and pans she used daily. Finally an analysis brought to light the fact that she had no conception of two dimensional perspective. Drs. Brigham and Healy[34] have warned against hastily assuming that pictorial material is universal; the former points out that pictures may be just as specialized as words, and that therefore in them as well as in verbal tests, we run up against "nativism"—that is, content adapted to a particular culture. Mead ('28) discussing the difference between primitive modes of representation and ours, indicates that Samoans could not grasp our conventional methods of indicating night and day, the rays of the sun and other natural phenomena. Miss B. M. Minogue[11] suggests that the Porteus Maze might be unfair to savages due to the dazzling effect of black and white on vision unaccustomed to it.

The language element enters into almost every kind of test, and is by no means eliminated in non-verbal tests. V. A. Jones ('27) says: "The fact that a test is composed exclusively of non-verbal material is no proof that it has merit as a non-verbal test" (p. 207). This is borne out by C. Kirkpatrick ('26), Estabrooks ('28a) and Koch and Simmons ('26). Walters ('24) after testing children thirteen years of age, who came from foreign-speaking homes, found that there was a language handicap of from six to eight months of mental age on the Binet Scale. (See also next page.) Even in the simplest performance test, it is not easy to analyze the possible learned elements; "nativism" and "experiencial material", so easy to recognize in language tests, disguise themselves in performance tests, so that an item which seems universal to us is actually limited to our environment because of some subtle method of portrayal or actual content. Davenport[11] points out that a jungle native and a man who shaves every morning would react very differently to the mirror-drawing test; and Kent, in a letter to the writers, criticized an item in her own Sentence Completion test, because it assumed that the number of sides and corners of a cube are common knowledge; as she herself points out, this presumption would immediately limit such a test as the Cube Construction to those groups having had at least the equivalent of kindergarten training. In the case of testing American Indians, Garth, who has worked with them perhaps more than anyone else, says ('25):[35]

> "One has but to visit the Indian in his hogan, his tepee or pueblo hut, or to see him in his community life, to be convinced of the meagreness of these as any sort of adequate preparation for the ways and customs of civilization .... Because of differences in social status and temperament, we cannot conclude that our results are true and final measures of Indian children."

Fitzgerald and Ludeman ('26) tell about a group of Indian children in a government school who were given the National Intelligence test. One of the questions was "Why should all parents send their children to school?" The "multiple choice" answers were: "(a) because it keeps them out of mischief; (b) it prepares them for adult life; (c) they are too young to work." The authors point out that answer "b" is meaningless to an Indian child who can see no relation between what he learns at school and the kind of life his elders lead on the reservation.

The above examples were chosen at random in order to illustrate how certain elements in different cultures — namely art-forms and social customs — can affect test results. Since almost everything in the physical and mental environment is included under the term "culture" (used in the anthropological sense), it is obviously hopeless to treat the topic of culture differences adequately. In accordance with the principle of offering illustrations to the reader instead of complete summaries, only the factors of language and socio-economic status will be described, because their influence on mental testing is perhaps most prominent.[36]

Language Differences

The study of the connection between articulate speech and thought forms one of the chief branches of anthropology, and the various proofs that have been brought forward to show how language structure influences ideas should vitally concern those who make mental tests.[37] Variations in the emotional tone of words, obvious even among European tongues, must surely be tremendous in the case of entirely different languages, such as the Nootka (Indian) and the Chinese. Boas ('28) notes that differences in family relations are reflected in the actual vocabulary; and elsewhere, speaking of mathematical ability, cites the case of tribes whose language groups numbers by fives, so that when higher units are wanted, four groups of five are combined (fingers and toes) with the result that counting is automatically done by twenties, in contrast to our tens. It is easy to see how our intelligence tests, containing as they do many mathematical elements, would be inapplicable to tribes with incomplete numerical systems. Goldenweiser ('22) is careful to state that such a deficiency is not due to lack of innate capacity, but represents a "peculiarity of the civilizational 'setting' ", meaning that where ideas of property and mediums of exchange are relatively unknown, progress in numerical computations is apt to be slow. In this connection, Mead[11] points out the impossibility of grading people who cannot count according to our numerical system.

As in the case of the speed factor, opinion is divided regarding the significance of linguistic ability for intelligence. Stockton says ('22):

> "When the mechanical use of language can be guarded against, language ability becomes one of the best evidences of ability to work on the symbol level in contrast to the perceptual."

Terman ('16, p. 230) maintains that vocabulary is the best single test of general intelligence. (See also page 65 of this volume). Snedden ('27) after reviewing the statistical evidence obtainable from the Alpha, Otis and Terman intelligence tests, surmises that "in a population having an S. D. of about 30 months mental age, a perfectly reliable vocabulary test would correlate with a perfectly reliable test of general intelligence about .88" (p. 34). In corroboration of this is the suggestion made by Bere ('24) and Goodenough ('26) that language difficulty is itself a symptom of low mental capacity. (The latter admits, however, that the length of time the particular national group has lived in America must be considered, and Klineberg ('28) adds to this "cultural patriotism", which may make one group prefer to retain their own language.) Drs. F. L. Wells,[11] Earle,[11] Tallman[11] and Bronner[11] are further upholders of the belief that language ability is closely related to intelligence in the sense that good language tests correlate closely with good intelligence tests. Burks and Kelley ('28) do not think it legitimate to infer "that language handicap accounts for the low scores of foreign children in verbal tests", when verbal and non-verbal test results on foreign children are compared. (See also Stockton '22 and Jones '27). On the other hand, many workers have argued that linguistic differences render verbal test results more or less invalid.[38] Pintner and Paterson, discussing their performance scale, point out that language ability is not always associated with intelligence, and Brown ('22) found that foreign children sometimes scored six to eight months higher if tested in their own language. A middle-of-the-road position is taken by Kirkpatrick ('26) who, finding differences between two groups of Finns and Italians, concludes that the linguistic handicap under which the latter labored makes the differences less than they appear, but "fails to disprove their existence". (p. 104. See also Appendix I, No. 66) Hirsch ('26) concludes his study of natio-racial differences with the statement:

> "At least thirty per cent of the mental differences between the English and the combined non-English speaking groups as tested in the Army was due to a language (and perhaps in part emotional) handicap, but the balance of the differences was due to a genuine superiority of the former." p. 350.

Obviously more work is needed, in regard to the relation of language to intelligence, in practical field work with performance tests, and in the re-testing of immigrants who have had opportunity to learn English. Ever since the Tower of Babel, language differences have been a stumbing block in the path of the internationally or inter-racially minded. Perhaps a verbal test in Esperanto will some day be found feasible, but that is a long way off. Non-verbal tests may offer a partial solution, but at the present stage of test development, their findings should probably be supplemented by verbal tests. (For further discussion of this subject, see Part Three, Chapter One.)

"Socio-economic status" as a factor in general culture, presupposes a stratification of classes according to definite social and financial criteria of our own. For this reason, such status is practically impossible to assess in a primitive community.

Social and Economic Differences

Blackwood ('27) shows how the "prerogatives of religious observance, marriage, and other relationships" so complicate the social organization of many tribes that even if it were well enough understood to be analyzed, it could never be used as a basis for comparing one group with another. In America, the socio-economic factor involves peculiar difficulties for inter-racial testing, because of the Negro problem and immigration. There is no denying that the coloured people still suffer from grave social disabilities, while in the case of European immigrants, complications arise from the fact that the economic attraction of this country was different for northern and southern Europeans. Each immigrant of the laboring class coming through Ellis Island tends to be thrust into the social and financial pattern which his racial group has created. An Italian, for instance, will in most cases step into a lower economic level than a Russian Jew, and, unless he has unusual abilities, will remain there. (See Neifeld '26, Boody '26, Klineberg '28, Pintner '29, and the discussion of sampling difficulties, page 37 of this volume.)

Attempts to equalize social conditions for the purpose of inter-racial testing have been made by Mayo ('13), Phillips ('14), Pyle ('15), Derrick ('20), Waugh ('20) and Reinhardt ('27). Arlitt ('21) concludes that race norms are invalid which do not take social status into account, since she finds that the differences between children of the same race but of different social class are greater that race differences. Nutting ('26), speaking of the Negro's inferiority complex, says that mental ability is more individual than racial, and that social environment profoundly affects it. Mead ('25) also raises the question whether social status is particularly influencial with groups suffering from race discrimination, and Peterson ('28), speaking of the difference between Negro and white "sets for efficiency" says:

> "Adult activities are so deeply set in vocational habits and attitudes that it is hard to be sure of the degree of innateness represented in any results obtained."

In this connection, H. H. Long[29] asks whether we are justified in drawing conclusions as to the intelligence of children (i.e. Negroes in Washington, D.C.) who are tested with a scale already standardised on a group which has had far greater cultural opportunities than the Negroes. (See page 48.) Blackwood stresses the difficulty of estimating the effect of racial mixture on an individual's social standing: sometimes it is an advantage, but more often it means that the offspring are branded as half-breeds and as such are cast out from both parent races. Even apart from the question of miscegenation, involving as it does complex problems both of heredity and environment, the fact remains that we know very little about "socio-economic status"— how large a part innate ability plays in causing its graded hierarchies, and what subtle effect it may have on achievement. Certainly it is a powerful factor, and one whose interaction with organic tendencies still remains obscure.

This review of the major hazards encountered when attempts are made to compare races is admittedly incomplete. It omits many different things in which races might vary and which undoubtedly affect test scores—factors such as law, religion, schooling, mental growth and "rapport". Obviously our typical intelligence tests, built for American School children, do not (and should not) take all these factors into account. This being so, such tests are not valid instruments for comparing races or any other diverse groups.[40] There are three ways in which tests might be used as tools of comparison: (1) by testing infants (2) by evolving some sort of universal yardstick, fashioned of material equally familiar or equally novel to each subject, and (3) — a more efficient but more limited edition of (2) — by scaling on two groups test material which is common to both, and using the resulting two sets of items on each group. The first method has never been undertaken on a large scale (See McGraw, p. 30) and will not be discussed here. The other two methods will be briefly described under the titles "universal testing" and "intra-group scaling".

UNIVERSAL TESTING

"Universal testing" has once or twice been attempted, but since, as we shall see, the majority of psychologists think that "material equally novel or familiar to man" does not exist, such testing can never be perfect. (See Appendix I, No. 52). The principle underlying it — namely, that a tool of comparison must measure the amount of a homogeneous capacity,— means that only those groups having enough common experience to make a measuring unit possible can be compared.

It might be well to apply this principle and see how limiting it is to a "universal" testing program, which connotes testing highly civilized people and primitive tribes with one instrument. Here "experience equally novel or familiar" is restricted to simple sensory processes. It is a moot question as to how far the sensori-motor measurements developed in Europe and America are truly objective. Assuming that they are — i. e. that they do not merely represent our own cultural elaborations — they provide a real tool of comparison with other civilizations. For instance, W. H. R. Rivers, who was in charge of the psychological testing on the Cambridge Expedition to the Torres Straits (1898), used the experimental methods fresh from the psychological laboratories of Europe to measure visual and auditory acuity, colour vision, hearing, smell, taste, space perception, cutaneous sensations and reaction time. He states a need for "sensorial measurements capable of universal application", and intimates that stars and rainbows could be used in testing visual acuity.[41] He comments favorably on the fact that, because "Martin's Pseudoptics" amused the natives, their cooperation was enlisted and they were presumably tested accurately. In such visual illusions as the "Man and Boy", he suggests that drawings of natives be substituted for the conventional pictures of Europeans in street clothes, for then the subject's attention would not be diverted from the actual test situation by the strangeness of what he sees. (For later accounts of testing sensory acuity among primitive peoples, see Ranke '06, Bruner '08, Woodworth '10 and Myers' 11). With but few exceptions, it can be said that the senses and elementary reactions of primitive peoples are found to be strictly comparable to those of civilized white men. Therefore, until more is known about the correlation between mental and motor reactions, our judgment of primitive intelligence must be suspended. Above all, degree of intelligence cannot be inferred from cultural achievement, because all such deductions are necessarily based on one particular standard. Anthropologists have analyzed the life of primitive man in many parts of the world, but few have succeeded in creating more than "a sympathetic account of externals, perhaps supported by native explanations more or less moulded by the intention of the investigator".[42]

Today the interest in universal tests extends beyond sensori-motor tests to the measurement of intelligence. Dodd ('26) has constructed a scale which aims at testing ability "through training common to humanity". The elements of world wide common experience, if it exists, could be listed roughly as follows: sensori-discriminations such as shape, roughness, dryness, heat and cold; the human body, whole or in parts, portrayed in different activities; childhood and old age; basic emotions, such as elation and anger; animals such as generic birds, insects and dogs; natural objects and scenes—sun, moon and stars, shadows, reflections, stones, fire and rain. On more debatable ground are suggestions such as artificial objects (knives, bowls, staves and beads), basic geometric figures, and number combinations up to five.[43] In the matter of actual test devices, Dodd points out that one is limited to fundamental reaction methods, such as pointing, pushing, twisting and grasping; to these Dr. C. B. Davenport[11] adds marking and scratching movements in general.

Assuming, then, that "universal" human experience exists beyond sense discriminations, the question of the significance of universal testing resolves itself to this: Can the higher mental processes be measured by test devices whose contents belong in the above universal categories, or do they have to be tested with complex symbols, peculiar to specific cultures? Rousseau said, "The accent and tone of voice vary between one nation and another, and signify different feelings among different peoples." If this is true of the voice, it is probably true of gestures also, in which case even pantomime directions cannot be standardized. How much more difficult will it be to find test *material* that does not have different meanings among different peoples!

Opinions of psychologists and anthropologists on universal testing

One of the present writers asked a number of psychologists and anthropologists whether they thought it possible or useful to evolve universal tests. Some thought, with reservations, that it was;[44] others surmised that once the material was pared down to a point where it was "universal", little of diagnostic value would remain;[45] while the majority were skeptical about the possibility of ever establishing a universal standard, because they felt that test elements peculiar to one type of experience could not be eliminated.[46] As was to be expected, the ranks of anthropology were arrayed against the project of testing civilized and primitive peoples with one instrument, their method of objective inquiry being to enter each culture on its own basis. Mead[11] raised practical objections against it, such as the fact that homogeneous primitive groups do not exist in great enough numbers to obtain statistically valid results; second, that there can be no developmental age studies among primitive peoples, since no birth records are kept; and third, there is no way of telling whether a native is putting forth his best effort, because nine times out of ten he does not feel a spark of interest in the test. W. H. R. Rivers himself remarked about the state of "abject fear" in which the native approaches the test situation, thus making it impossible to judge of his real abilities; and Bartlett ('23) emphasizes the importance of obtaining the primitive subject's "active cooperation".

A few quotations will suffice to illustrate the prevailing views on the subject of universal testing in anthropological circles:

"In none of the tests is individual experience eliminated, and I doubt if it can be done."— Boas '28, p. 55.

"No intelligence tests have been devised which bring within the same realm for comparison on an equitable basis the civilized man and the savage."—Wallis '26, p. 321.

"The application to primitive peoples of tests we use among ourselves involves two assumptions: first, that the nature of intelligence is essentially the same throughout; and second, that reactions which are symptomatic in one set of conditions, will be equally symptomatic in another. (p 4) .... To quote Thorndike, we are .... 'measuring differences between groups which are distinct to an unknown degree in traits which are influenced by training to an unknown degree'— he might have added, 'by means of instruments which are accurate to an unknown degree.'" .... Blackwood '27, p. 15.

Thus the judgment of experts, both psychological and anthropological, is, on the whole, unfavorable to universal testing. It would be surprising if it were not, as any world-wide standard would presuppose agreement on, first, the nature of intelligence and, second, the best means of measuring it. The first is still in the field of debate, as we have seen, and as for the second, no two authorities agree as to fundamental test forms.[47] Some unity of opinion is urgently needed, even though the probability is that, since many tests are necessary to measure even one mental function, no single basic item exists. A number of mental testers are engaged in the statistical process called "item analysis", which is, essentially, ascertaining which elements in a test are valid, that is, which items discriminate best between subjects who are good and poor according to certain defined criteria.[48] Once this movement gains momentum, much valuable information will be obtained and tests will be more reliable as well as more valid than they are now.

Suggestions for universal tests

Although there is no agreement as to basic items, some suggestions for world wide tests were made by those psychologists who thought it might be useful to test people of different cultures. One[49] advocated translating the Binet Scale outright, with alterations only when idioms or the "experiential element" demand them; others[50] suggested that in the case of testing primitive peoples, the nearest approach to "fairness" would be to take native "intelligence tests", such as initiation feats, string games, and sign languages, and then use them alike on primitive and civilized groups. Some[51] supported pictorial and drawing tests as the most universal, provided the natives could manipulate a pencil; and others[52] urged formboards. Again, tests which tap mathematical ability in some concrete form were considered best,[53] and the same holds true for linguistic tests composed, let us say, of nonsense syllables. Finally, many authorities[54] asserted that tests which involve the learning process offered the most hopeful line of advance, because the effects of practice could be controlled all over the world, regardless of culture.

This last proposal has been so widely supported in psychological literature, that we are justified in giving it more than passing attention. Since learning capacity is closely connected with intelligence, any test technique which makes use of practice or training has a direct bearing on the measurement of innate ability. If individual differences are eliminated by training, they are presumably due mostly to circumstances, but if they persist even after equalized training, they may be attributed more to nature than to nurture.[55] Starch ('20) says:

Tests of the learning process

"All experimental results point in the direction that practice does not equalize abilities; in fact, equal practice tends to increase differences in achievement and skill rather than to decrease them. The more gifted individuals profit more, both relatively and absolutely, than the less gifted." p. 91. (See also page 32 of this volume.)

The influence of practice, then, is of fundamental importance when interpreting test results, even though all human traits are not subject to training in the same degree. Slocombe ('26) calculates the "index of intellective saturation" for many retestings with several kinds of tests. He concludes that, since this index is high, at least 25 per cent of testing time should be given over to "fore-practice" exercises. Not a few tests use such exercises — e.g the National Intelligence Test, O'Rourke Non-language Tests and the Dodd International Scale. Orleans[11] proposes a scale of this type — a series of lessons with tests based on them, whose content would be universal. In a letter to the writers he says: "If the individuals constructing the material are sufficiently ingenious .... it would be possible to develop a test .... ranging from the kindergarten up and involving different types of mental functions .... which may be usable anywhere." Blackwood ('27) referring to the Dodd Scale, says: "The plan of practising the subject before the real test is likely to serve good purposes among primitive peoples, most of whom will be entirely unfamiliar with any form of test situation." She adds that the scores of the Indians she was testing increased after the practice pages, and that familiarity with what was expected of them "induced a confidence which probably made their performance after practice a closer approximation of their real ability." In 1904 at the St. Louis Fair Woodworth gave members of various races repeated tests with a formboard, and found that the initial superiority of the whites, probably due to familiarity with the test situation, was not maintained. Dodd ('26) adds to a list of improvements needed in universal testing the importance of practising foreign subjects till they are familiar with the test situation. (See page 47 of this volume.) In constructing the association test in Form B of his scale he aimed to control the position and gradation of difficulty so that the laws of learning and memory could be studied universally. Ruger[11] thinks that if the transfer principle (i.e. learning) were utilized, people of different environments could be tested fairly. The important thing is to start with simple situations, composed of elements familiar to the subject, and then increase their complexity. Of like mind is Graham[56], who writes that if tests are to be of value in inter-racial testing they should begin with problems whose solution is obvious, and then proceed to related examples of gradually increasing difficulty.

Existing Universal Scales

Leaving the realm of theory and opinion in order to discover whether any existing tests claim to be universally applicable, we find only three which have been constructed and used for this special purpose. Other tests, such as the Porteus Maze, the Disc Transfer Test, the Bead Test, and the Color-Form Substitution Test, which might be thus used, are described later. (See pages 83, 84, 183 and 200; also Appendix I No. 63.) The first is a scale of individual performance tests which ranges from low to very superior levels of intelligence. (Squires '26). This can be administered with pantomime instructions, and is therefore a true non-language test. The author is careful to say, however, that although it is independent of language so far as instructions and responses are concerned, it does not necessarily avoid the language element in the subject's mental operations. The test material is limited to a four-bar xylophone, wooden cubes, dot patterns, "universal" picture material, and the like. Unfortunately, the scale has only been tried out on a group of fifty "superior" adults (Princeton students) and fifty feebleminded male inmates of a State institution (Vineland). The author suggests that it be used in inter-racial testing, and as this is the only individual performance scale that has a claim to universality, its employment with another race would yield not only interesting but valuable results — if not in the way of racial differences, at least as regards testing technique.[57]

Squires

Peterson

The second test which, although it does not dispense with language, has been built so that it shall not "exceed the range of learned skill or information in any one group" is Peterson's Rational Learning Test ('23, and, with Lanier, '29). Here the subject must associate certain numbers with certain letters which he hears orally; the faster he discovers the principle underlying different combinations, the more successfully or "rationally" will he perform. Peterson points out that although this test "taxes thinking activities to the extreme", it stimulates each subject alike, and therefore does not favor the fast ones. Because it separates speed from accuracy and number of responses, this test is able to evaluate different races on three factors independently.[58] It has been used on whites and Negroes both in Nashville, Tennessee and in New York. The differences in the first locality were significantly smaller than those obtained when Myers Mental Measure or Dodd's International Scale were used on the same group.

Dodd

The third universal test has just been mentioned (Dodd '26). It is a group scale which makes use of a rotator device whereby the subject may record his choice of a picture without using a pencil. Affixed to each page are many cross-shaped pieces of cardboard, each of which must be turned till one of the four pictures on the four arms appears in such a position as to fulfill a special function, such as analogy completion. This test also can be given with pantomime instructions, and these have been carefully standardized by means of cinematographic studies and micromotion analyses. The author ably discusses the problems involved in the construction of a universal test, and lays special emphasis on the importance of discovering world wide local criteria of success other than scholastic achievement. The Dodd Scale has been used on feeble-minded persons, college students, orphans, Mexicans, Negroes, Indians and Hindus. The results of the last experiment (conducted by Dr. C. Herbert Rice) showed that the average of 200 Hindu children was lower than that of American subjects. Before interpreting this as meaning that the Hindus are inferior, other explanations must first be carefully considered. For instance, a more leisurely civilization than ours does not train its peoples to work according to a time limit, and thus the children's attitude may have been a striving for accuracy rather than for speed.[58] Even more important, the tasks, although composed of universal elements, present situations calling for mental "sets"— mazes, picture completions, and puzzle elements in general — which Hindu children are not apt to have encountered unless they have gone to a mission kindergarten. As a result of these considerations, Dodd suggests that the following improvements be made, before racial differences in intelligence can be compared on the basis of test performance: (1) better test items should be devised; (2) different time limits should be established; (3) foreign subjects should be practised until they are familiar with the test situation; and (4) the tests themselves must be adjusted until they give high correlations with criteria of success in the local environments. Despite the faults implied by these needs, Dodd concludes:

> "We believe that the normal distribution given by these tests in India, and the correlations they yield with local criteria in completely different civilizations, shows that an international mental measuring instrument is possible, and may be refined by research until it becomes a very useful instrument in many countries."—'26, p. 87.

Similarly, Blackwood ('27) although she did not find reliable differences between the scores of 413 Indian and 200 Mexican children on the Dodd Scale, concludes that "a beginning has been made in the construction of a test suitable for primitive conditions."

Thus, an "international measuring instrument" and a "test suitable for primitive conditions" are within the realm of possibility. Neither of these could be accurately called a "universal" scale, for neither implies that the same test should be useful in primitive and civilized conditions alike. We have seen that many psychologists think this is impossible for the following reason: even if all material which obviously belongs to a specific culture is excluded from a test, there will remain certain ineradicable features peculiar to any one method of representation, and these will penalize persons who are unused to that method.

INTRA-GROUP SCALING

One psychologist, Dr. David Shakow, wrote to the authors that although he shared the above objection, he thought it important to adapt tests to primitive peoples in order to learn about their mentality. Dodd himself asserts that a proper test for a primitive tribe should be constructed from the natives' viewpoint, after the maker has lived with them for a number of years.[59] It is a well known fact that the better suited a test is to the group on which it is used, the more accurate it is as a tool of differentiation. Dunlap ('27) says: "The usefulness of a test varies directly with its applicability to the class of persons for which it was devised", and Colvin ('22) bases the validity of an intelligence test on its being "within the common experiences" and appealing to the "common interest" of the group. For this reason the latter thinks that instead of the present general norms (which he considers on the whole of little value) there should be established norms for each particular group tested. A similar conclusion was reached by Wood ('29) who tested Bulgarian, Turkish, Armenian and Greek women students with the Otis Self Administering test. She found, as was to be expected, large training and language handicaps; and this caused her to demand, if not special tests, at least separate norms for this group. Of like mind is Dr. Stenquist, Director of Educational Research in Baltimore, who thinks that the colored children in the public schools should have their own tests and standards.[11] (See also Petersen '28 on the uselessness of white norms for race comparisons.)

Pintner ('26) shows how the group used for standardizing a test determines the background against which the ability of a new individual is to be judged: since a specific reaction can only be used as a test of intelligence if the background is equally novel or familiar to each testee, it follows that the new individual must approximate the standardization group, and a test built and standardized for one group will not do for another. Writing on this general subject, Kent[60] has given the opinion that if certain tests were standardized for different classes of subjects, their range of applicability would be greatly increased. Shimberg ('29) has actually examined the differentiation of norms according to the sex, education, locality and racial composition of rural and urban children. She concludes that the original scaling of a test usually favors one group to the exclusion of the other. (See page 51.) Gordon ('23) tested canal boat children in England, and found very little mental development on the intellectual side. He attributes this mostly to lack of schooling, and says:

Principle of Multiple Norms

> "How far there has been a similar lack of development among these children in connection with problems touching their own special environment, it is difficult to say, and, *without tests especially devised and standardized for children in such surroundings, impossible to measure*". (Italics ours)

Indeed, as far back as 1919 Pressey and Cole criticized the practice of testing one group (adults) with tests devised for another (highschool pupils). They urged that special tests should be evolved, because ordinary test material has little connection with the problems by which adult intelligence is actually proved.

In the case of racial testing, efforts to escape this basic fallacy—this comparing of two groups with instruments adapted to only one of them—have generally taken the form of adapting our tests to different peoples for the purpose of classification within the group rather than for comparison with America. One of the best examples of this is Liu's "Non-verbal Tests for Use in China" ('22). Here a Chinese psychologist, trained in America, selected the most useful elements from well-known American non-verbal tests, and prepared alternative forms, because, as he says:

Tests adapted to other groups

> "The uncivilized Miaotze boy in Yunnan could not be expected to answer questions on automobiles or airplanes, and the New York boy, raised in the Bronx, could not be expected to answer intelligently questions on rice growing . . . ." p. 14.

Walters ('27) reports another set of tests created especially for use in Porto Rico, because American tests were found to be unsuitable. The scale which he describes is built on an American model but is based on Porto Rican environment. Spanish is used instead of English, and questions are asked about sugar cane instead of potatoes. A third example of tests built for a different people is Helma's experiment with test material based on experiences common to three Indian tribes. With this she tested both whites and Indians, and, as was to be expected (since only the Indian test was used), the items that gave best results for the Indians were least adapted to the whites. (Ref. Unpublished Thesis, University of Kansas.) This brings to mind the trenchant observation made by W. D. Wallis ('26):

> "Devise tests which suit the intelligence and the interest of the native Zulu, and he will do better than the average white. The relative standing depends considerably upon who devises the tests and upon the criteria imposed".[61]

Adapted Binets

There also exists a group of studies in which the Binet Scale has been adapted for use in different countries. Martin ('15) reports an experiment on thirteen Kafir school children in Africa. Here the Binet tests for younger ages were found to be the best, for above the seventh year too many items had to be omitted — (e.g. dates, rhymes, coinage, etc.). The color distinctions created difficulty also, for green and blue are not easily discriminated by the Kafirs, and no word for "yellow" exists. (See Appendix I, No. 43.) Loades and Rich ('17) applied the Binet under really primitive conditions in Natal, South Africa. They adapted the Goddard Revision to the requirements, translating and substituting items as they saw fit. Thus, instead of the key, knife and penny in test two, age four, they used a native pot, beads, and a loin cloth. The pictures to be described were changed because "the regular picture implied a cultural status which even the most educated Zulus have not attained". In 1921 Yeung, testing Japanese and Chinese children on our west coast, altered the Binet so that the language difficulty would be lessened, while in the same year Kubo translated it outright in order to test 1200 children in Tokyo. Porteus and Babcock ('26), working with Hawaiian, Japanese, Chinese and Portuguese children in Hawaii, separated the Binet items into tests of native ability and tests most influenced by training, and scored accordingly. Finally, a Hindustani Binet with verbal and non-verbal sections has been constructed by Dr. C. Herbert Rice and used to compare Brahmins, Outcasts, Moslems and Christians in India ('29).

Ideal Inter-group Testing

Because the tests used in these studies are more or less different from their American prototypes, the "reactions symptomatic of one set of conditions" will not be "equally symptomatic in another". In other words, these studies are to be described as "*intra*-racial" not "*inter*-racial." If some sort of scale could be devised which would assemble enough test elements so that the environment of each candidate would be fairly represented, and through all of which would run the same testing technique—for example, the measurement of speed or learning—then performance on its different sections might be more nearly comparable. Thorndike's principle of making a scale so embracing that it would sample many fields of interest so that when subjects with different backgrounds are tested, each would find something familiar and no one would complete the test,"[62] lends support to such a suggestion. Ruger[11] has also stressed the need for scales composed of many different materials, but with the same organization, so that performances on the different sections can be compared. A concrete suggestion is found in the study by Carreon ('26), who urges that certain items in the Haggerty be modified so that they would have "the same significance and relative difficulty" for the Filipino child as for Americans. Thus there would ultimately be obtained Philipine scales with standards both for local purposes and for comparison with America. (See also Walcott '20.) Testing of this sort—probably the ideal way to compare groups—can only occur when the different peoples of the world understand each other's cultures and criteria of success. Eventually we may realize that a Zulu who gets a high score on his test is just as intelligent as an American with an I. Q. of 150—that is, if the Zulu can defend his own culture against the encroachments of western civilization. At present, however, we must cling to the principle that a common measuring unit is the only basis on which similarities or dissimiliarities may be revealed.

Intra-group Scaling

So far, as we have seen, the tests that can be called true tools of comparison between races are few. About half a dozen studies[63] have been completed in which an approach is made towards obtaining scores correlating with criteria of mental and social success in different environments. Actually the environments can never be very different, for each must contain material common to the other, and must also have provided training in the basic "testing idea." "Universal testing" is thus limited; and if tests are to be used as tools of comparison and at the same time be really efficient tools within each group, some modified scheme of testing is needed.

We have seen that the most valid tests are those best adapted to their own homogeneous groups.[64] Indeed, as regards racial studies, many psychologists (viz. Drs. V. A. Jones,[11] A. Bronner,[11] Rowe, '14 and Peterson '28) think that *intra-racial* testing is more important practically than are racial comparisons, because the former can aid in the adjustment of the various social and economic groups within a race. This in itself, as a matter of fact, would be inter-group testing, since social differences are often much larger than racial or national ones (Arlitt '21 and Scott '30). Thus we cannot let ourselves consider that group comparisons are odious; the necessity for them keeps confronting us, especially if we are interested in diagnosing nature and nurture.

A modified scheme of comparative measurement, which would take advantage of the principle that a test functions most usefully when adapted to a particular group, might be worked out somewhat as follows: In the case of comparing two specific groups, a list of items could be drawn up which would not necessarily be "universal" but which would omit all elements exclusively local to each group. This would assuredly be difficult, since every group comparison involves different homogeneities and heterogeneities; it could be done, however. This preliminary test could then be scaled and standardized on each group, with the result that different items would automatically attain separate difficulty values in each, and two scales would emerge. Each scale could then be used on both groups, and the resulting two pairs of scores would yield two pairs of differences from which superiorities or equalities could be computed. If one group did better on the other group's test in comparison with its own than vice versa, it could be regarded as superior.

Shimberg ('29) has done almost this very thing with urban and rural children. Maintaining that no one can tell *a priori* what test questions will be "fair" to any group, she had two preliminary information tests of 80 questions each prepared—one for a group of city children and the other for a matched group of country children. In each case, the children's teachers made up questions they thought were fair to their group, but neither test contained purely urban or rural material; in fact, thirty-seven of the questions were identical. She then T-scaled and standardized each test on its own group, after which she tried it out on the other. The result was that the rural children did *less poorly* on the city test than the city children did on the rural one. The conclusion is, quite naturally, that the mental differences usually found between urban and rural children are not due to innate differences but are a "function of the tools of measurement". In other words, the "city tests do not fit our country cousins".

Conclusion

We can say in conclusion, that "intra-group scaling" such as has been described above may shed light on the basic dilemma of group comparisons. A person instead of being compared with someone from a different group who may be more familiar with the test, will be measured with a scale in which the parts that are hard for him will be balanced by corresponding difficulties for the other, even though each actual test item is not equally familiar to both. In other words, a compromise will have been effected between tests as tools of comparison—universal tests—and as tools of classification—specialized tests. Such a scale would be less efficient within each group than a test using highly local material; and it could not be used to compare more than the specific groups for which it was built and scaled. It stands to reason that it would be unprofitable to compare peoples differing too widely.

In the last analysis, as Dr. Wissler[11] has pointed out, to measure a person is nothing more than to compare him with others who are more or less like him—i. e. in his group; and the only way to draw true conclusions as to what a test tests is to consider three factors together: the test, the subject tested (including his milieu) and his performance. (See Appendix I, No. 3.) Perhaps the ultimate basis of comparison between the peoples of the world will be the differences between their criteria of achievement, but that is impossible until universal standards are universally agreed upon. Then, perhaps, specific groups all over the world, differing racially or culturally, will be supplied with tests well suited to each one and at the same time yielding statistically comparable scores. Until that time, however, some more limited process such at the "intra-group scaling" suggested above appears to be the most practical way of making comparisons.

Testing of this last sort would measure innate ability by sorting out the effects of different environments. For example, once two socio-economic "populations" had been delimited, infants from group "a" might be transplanted into group "b" and there reared and tested; the scores made by sample "a", when compared with both the "a" and "b" norms, would supply information about the two environments. Hereditary features would then stand out, and any one of the manifold varieties of environment—be it social class, school, vocation, or any other—could be tracked down, so that its potency "to add to or to detract from human endowment"[65] might at last be ascertained.

BIBLIOGRAPHY

References cited in Part II, Chapter II.
Studies of Mental Inheritance.

(See Burks '28 b pp. 252–268, 282–291 for annotated supplements to this Bibliography.)

Antipoff, H. '28. L'évolution et la variabilité des fonctions motrices. Arch. de Psychol., 21, 1–54.

Arlitt, A. H. '21. On the need for caution in establishing race norms. J. Appl. Psychol., 5, 179–183.

Averill, L. A. and Mueller, A. D. '25. Physical and mental measurements of fraternal twins. Ped. Sem., 32, 612–627.

Baldwin, B. T. '28. Heredity and environment—or capacity and training? J. Ed. Psychol., 19, 405–409.

Banker, H. J. '28. Genealogical correlations of student ability. J. Hered., 19, 503–508.

Bartlett, F. C. '28. Psychology and primitive culture. New York. Pp. 294.

Bere, M. '24. A comparative study of the mental capacity of children of foreign parentage. T. C. Contrib. to Educ., 154, Pp. 105.

Bernstein, E. '24. Quickness and intelligence; an enquiry concerning the existence of a general speed factor. Brit. J. Psychol. Monog. Suppl. Cambridge Univ. Press, Pp. 55.

Bickersteth, M. E. '19. The application of mental tests to children of various ages. Brit. J. Psychol., 9, 23–73.

Blackwood, B. '27. A study of mental testing in relation to anthropology. Ment, Meas. Monog., 4, Pp. 120.

Boas, F. '11. The mind of primitive man. New York: MacMillan Co. Pp. 294.

—'21. The Negro in America, Yale Rev. 384–395.

—'27. Fallacies of racial inferiority. Curr. Hist. Feb. 681–682.

—'28. Anthropology and modern life. New York: W. W. Norton & Co., Inc. Pp. 246.

Bond, H. N. '26. Non-intellectual traits of Negro adults. J. Abn. & Soc. Psychol., 21, 267–276.

Boody, B. M. '26. Testing immigrant children. Baltimore: Williams & Wilkins, Pp. 163.

Brigham, C. C. '23. A study of American intelligence. Princeton, N. J.: Princeton Univ. Press, Pp. 194.

Brimhall, D. R. '22 & '23. Family resemblances among American men of science. Amer. Nat. 56, 504–547, 57, 137–152, 326–344.

Brown, G. L. '22. Intelligence as related to nationality. J. Educ. Res., 5, 324–327.

Brown, W. and Thomson, G. '21. The essentials of mental measurement. Cambridge: Cambridge Univ. Press, (3rd ed.) Pp. 216.

Bruner, F. G. '08. The hearing of primitive peoples. Arch. of Psychol., 11, Pp. 113.

Burks, S. '28. The relative influence of nature and nurture upon mental development; a comparative study of foster parent—foster child resemblance, and true parent—true child resemblance. 27th Yrbk. Nat'l Soc. Stud. Educ. Pt. I. 219–315.

—'28 b. A summary of the literature on the determiners of the intelligence quotient and the educational quotient. 27th Yrkb. Nat'l. Soc. Stud. Educ. Pt. II. 248–253.

—'29. Note on Prof. Freeman's discussion of the Stanford study of foster children. J. Educ. Psychol., 20, 98–101.

— and Kelley, T. L. '28. Statistical harards in nature–nurture investigations. 27th Yrbk. Nat'l. Soc. Stud. Educ., Pt. I. 8–38.

Burt, C. '12–'13. The inheritance of mental characters. Eug. Rev., 4, 168–200.

Candolle, A. de '73. Histoire des Savants depuis deux siclés. Geneva: H. Georg. Pp. 482.

Carreon, M. L. '26. Applicability of standard tests to the Filipinos. Yonkers: World Bk. Co. Pp. 175.

Cattell, J. M. '21. A statistical study of American men of science. New York, Pp. 27, 3rd Ed. Reprinted from Science, 24, 658–665, 699–707, 732–742.

Colvin, S. S. '22. Principles underlying the construction and use of intelligence tests. 21st Yrbk. Nat'l. Soc. Stud. Educ. 11–45.

Colvin, S. A. and Allen, R. D. '23. Mental tests and linquistic ability. J. Educ. Psychol. 14, 1–20.

Constable, F. C. '05. Poverty and hereditary genius; a criticism of Mr. Francis Galton's theory of hereditary genius. London: Fifield. Pp. 149.

Davenport, C. B. and Craytor, L. C. '23. Comparative social traits of various races. Second study. J. Appl. Psychol., 7, 127–132.

Derrick, S. M. '20. A comparative study of the intelligence of 75 white and 55 colored college students by the Stanford revision of the Binet-Simon Scale. J. Appl. Psychol., 4, 316–329.

Dodd, S. C. '26. International group mental tests. Princeton, N. J.: Prince. Univ. Store, Agents. (Mimeographed) Pp. 101, plus appendix.

Downey, J. E. '18. Standardized tests and mental inheritance. J. of Heredity, 9, 311–314.

Duff, J. F. and Thomson, G. H. '23. Social and geographical distribution of intelligence in Northumberland. Brit. J. Psychol., 14, 192–198.

Dugdale, R. L. '10. The Jukes. New York: Putnam (4th Ed.) Pp. 120.

Dunlap, K. '27. Social psychology. Baltimore: Williams & Wilkins, Pp. 261.

Elderton, E. M. '23. A summary of the present position with regard to the inheritance of intelligence. Biometrika, 14, 378–408.

Ellis, H. '04. A study of British genius. London. Pp. 300.

Estabrook, A. H. '16. The Jukes in 1915. Publ. Carnegie Institution of Washington. No. 240. Pp. 85.

Estabrooks, G. H. '28 a. A proposed technique for the investigation of racial differences in intelligence. Am. Nat. 62, 76–86.

—'28 b. That question of racial inferiority. Amer. Anthrop., 30, 470-475.

Feingold, G. A. '24. Intelligence of the first generation of immigrant groups. J. Educ. Psychol., 15, 65–82.

Ferguson, G. O. Jr. '16. The psychology of the Negro. Arch. of Psychol., 5, No. 36. Pp. 138.

Fisher, R. A. '18. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal. Soc. Edinburgh. 52, 399–446.

Fitzgerald, J. A. and Ludeman, W. W. '26. The intelligence of Indian children. J. Comp. Psychol., 6, 319–328.

Freeman, F. N. '26. Mental tests: their history, principles and applications. Boston: Houghton Mifflin Co. Pp. 503.

—'28. An evaluation of evidence in Part I of the Yearbook and its bearing on the interpretation of intelligence tests. J. Ed. Psychol., 19, 374–380.

—'29. The resemblance of identical and fraternal twins in a number of traits. Proceedings, 37th annual meeting American Psychological Assn. Psychol. Bull., 26, Pp. 161.

—, Holzinger, K. J., Mitchell, B. C. '28. The influence of environment on the intelligence, school achievement and conduct of foster children. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I. 103–217.

Galton, F. '83. Inquiries into human faculty and its development. New York: Macmillan Co. Pp. 387.

Garrison, K. C. '29. An investigation of some simple speed activities. J. Appl. Psychol., 13, 167–172.

Garth, T. R. '21a. White, Indian and Negro work curves. J. Appl. Psychol., 5, 14–25.

—'21b. The results of some tests on full and mixed blood Indians. J. Appl. Psychol., 5, 359–372.

—'22. A comparison of mental abilities of mixed and full blood Indians on a basis of education. Psychol. Rev., 29, 221–236.

—'23. A comparison of the intelligence of Mexican and mixed and full-blooded Indian children. Psychol. Rev., 30, 388–401.

—'25. The intelligence of full-blooded Indians. J. Appl. Psychol., 9, 382–389.

—'27. The intelligence of mixed blooded Indians. J. Appl. Psychol., 11, 268–275.

— and Barnard, M. A. '27. The will-temperament of Indians. J. Appl. Psychol., 11, 512–518.

— and Garrett, J. E. '28. A comparative study of the intelligence of Indians in the United States Indian Schools and in the Public or Common Schools. Sch. and Soc., 27, 178–184.

Gates, A. I. '21. The inheritance of mental traits. Psychol. Bull., 18, 358–365.

—'28. Observed facts and theoretical concepts. J. Educ. Psychol., 19, 381–388.

— and Taylor, G. A. '25. Experimental study of the nature of improvement resulting from practice in a mental function. J. Educ. Psychol., 16, 583–592.

Gaw, F. A. '25. A study of performance tests. Brit. J. Psychol. (Gen. Sect.) 15, 374–392.

Gesell, A. and Thompson, H. '29. Learning and growth in identical infant twins. Genet. Psychol. Monog., 6, 1–123.

Goddard, H. H. '13. The Kallikak family; a study in the heredity of feeblemindedness. New York: Macmillan Co. Pp. 121.

—'29. Hereditary mental aptitudes in man. Eugenics, 2, 1–7.

Goldenweiser, A. A. '22. Early civilization. An introduction to anthropology. New York: A. A. Knopf. Pp. 428.

Goodenough, F. L. '26 Racial differences in the intelligence of school children. J. Exp. Psychol. 9, 388–397.

Gordon, H. '23. Mental and scholastic tests among retarded children. Board of Education, Educational Pamphlets, 44. London. Pp. 92.

Gregg, J. E. '25. The comparison of races. Sci. Mo. 20, 248–254. Southern Workman, 54, 70-75.

Gun, W. T. J. '28. Studies in hereditary ability. London: George Allen & Unwin Ltd. Pp. 288.

Haddon, K. '11. Cats Cradles from many lands. London: Longmans Creen & Co. Pp. 95.

Haldane, J. B. S. '29. Scientific calvanism. What if science destroys free will? Harpers Magazine, 159, 551–558.

Hankins, F. H. '26. The racial basis of civilization. New York: A. A. Knopf, Pp. 384.

Herrick, D. S. '21. A comparison of Brahman and Panchama children of S. India, with each other and with American children by means of the Goddard formboard. J. Appl. Psychol., 5, 253–260.

Herskovits, M. J. '26. On the relation between Negro white mixture and standing in intelligence tests. Ped. Sem., 33, 30–42.

—'27. The Negro and intelligence tests. Hanover, New Hampshire, Sociological Press, Pp. 14.

Heymans, G. and Wiersma, E. '06–'07. Beitrage zur speciellen Psychologie auf Grundeiner Massenuntersuchung. Zeit. f. Psychol., 42, 81 and 258; 43, 321; 45, 1.

Highsmith, J. A. '25. The relation of rate of response to intelligence. Psychol. Monog., 34 (No. 3) 1–33.

Hildreth, G. '25. Resemblance of siblings in intelligence and achievement. T. C. Contrib. to Educ. 186. Pp. 65.

Hirsch, N. D. M. '26. A study of natio-racial mental differences. Genet. Psychol. Monog. Vol. 1, Nos. 3 and 4, Clark University. Pp. 167.

Holmes, S. J. '21. The trend of the race. New York: Harcourt Brace & Co. Pp. 396.

Holzinger, K. J. '29. The relative effect of nature and nurture influences on twin differences. J. Educ. Psychol. 20, 241–248.

Hooten, E. A. '26. Methods of racial analysis. Science, 63, 75–81.

Hsaio, H. H. '29. The mentality of the Chinese and Japanese. J. Appl. Psychol., 13, 19–31.

Hunter, W. S. and Sommermeier, E. '22. The relation of degree of Indian blood to score on the Otis Intelligence Test. J. Comp. Psychol., 2, 257-277.

Hurlock, E. B. '26. The suitability of the Downey will-temperament test as a test for children. J. Appl. Psychol., 10, 64–74.

Jennings, H. S. '25. Prometheus. New York: Dutton. Today and Tomorrow Series. Pp. 86.

—'30. The biological basis of human nature. New York: W. W. Norton & Co. Pp. 384.

Jones, H. E. '28. A first study of parent-child resemblance in intelligence. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. 1, 61–72.

Jones, V. A. '26. Effect of age and experience on tests of intelligence. Contrib. to Educ. (No. 203) Teachers College, Pp. 74.

—'27. A study of the non-verbal nature and validity of Myers mental measure. J. Educ. Res., 16, 203-209.

Kelley, T. L. '30. The inheritance of mental traits. Chap. 23 in "Psychologies of 1930". Worcester, Mass.: Clark University Press. 423–443.

Kemal, C. '28. Contribution a l'étude des tests de development moteur d'Ozeretzky. Arch. de Psychol., 21, 93–99.

Kirkpatrick, C. '26. Intelligence and immigration. Ment. Meas. Monog. 2. Baltimore: Williams & Wilkins, Pp. 127.

Klineberg, O. '28. An experimental study of speed and other factors in "racial" differences. Arch. of Psychol., 93, Pp. 111.

Koch, H. L. '28. A study of a pair of Siamese twins. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I. 75–81.

— and Simmons, R. '26. A study of the test performance of American, Mexican and Negro children. Psychol. Monog. Vol. 35, No. 5.

Kretschmer, E. '30. The breeding of the mental endowments of genius. Psychiat Quar. 4, 74–80.

Kubo, Y. '21. The revised and extended Binet–Simon tests applied to Japanese school children. Ped. Sem., 19, 187–194.

Lauterbach, C. E. '24. Studies in twin resemblance. Genetics, 10, 525–568.

Liu, H. C. '22. Non-verbal intelligence tests for use in China. T. C. Contrib. to Educ. Pp. 126.

Loades, H. R. and Rich, S. G. '17. Binet tests on South African natives, Zulus. Ped. Sem., 24, 372–380.

Macrone, I. D. '28. Preliminary results from the Porteus Maze tests applied to native school children. So. African J. Sci., 25, 481–484.

Marett, R. R. (no date.) Anthropology. Home Univ. Library, New York: Henry Holt & Co. Pp. 256.

Martin, A. L. '15. Experiment with Binet test upon African colored children, chiefly Kafirs. Vineland, N. J.: Training Sch. Bull., 12, 122–123.

Mayo, M. J. '13. The mental capacity of the American Negro. Arch. of Psychol. 28.

McComas, H. C. '14. The heredity of mental abilities. Psychol. Bull., 11, 379–383.

McFadden, J. H. and Dashiell, J. F. '23. Racial differences as measured by the Downey will-temperament test. J. Appl. Psychol., 5, 73–76.

McFarland, R. '30. An experimental study of the relationship between speed and mental ability. J. General Psychol., 3, 67–96.

Mead, M. '25. The methodology of racial testing. Amer. J. Soc., 31, 657–667.

—'27. Group intelligence tests and linguistic disability among Italian children. Sch. & Soc., 25, 465–468.

—'28. Coming of age in Samoa. New York, William Morrow & Co. Pp. 297.

Merriman, C. '24. The intellectual resemblance of twins. Psychol. Monog., 33, No. 5, Pp. 58.

Morgan, T. H. '29. The mechanism and laws of heredity. In Foundations of experimental psychology, 1–45. Worcester, Mass.: Clark Univ. Press.

Muller, H. J. '25. Mental traits and heredity. J. Hered., 16, 433–448.

Myers, C. S. '11. On the permanence of racial mental differences. Papers on inter-racial problems. Ed. Spiller. The World's Peace Foundation, 73–79.

Neifeld, M. R. '26. The race hypothesis. Amer. J. Soc., 32, 423–432.

Newman, H. H. '29. Mental and physical traits of identical twins reared apart. Twins "A" and "O". J. Hered., 20, 49–64, 153–166.

Nutting, C. C. '26. The mentality of "inferior races of man". Sch. & Soc. 24, 89–96.

Odum, H. W. '10. Social and mental traits of the Negro. Studies in Hist. Econ. and Public Law. Columbia Univ. 37. Pp. 303.

Oldham, J. H. (no date). Christianity and the race problem. New York, Doran. Pp. 280.

Paschal, F. C. and Sullivan, L. R. '25. Racial differences in the mental and physical development of Mexican children. Comp. Psychol. Monog., 3, No. 14, Pp. 76.

Peak, H. and Boring, E. G. '26. The factor of speed in intelligence. J. Exper. Psychol., 9, 71–94.

Pearson, K. '01. On the inheritance of the mental characters in man. Proc. Royal Society, London, 69, 153–155.

—'03. On the inheritance of the mental and moral characters in man, and its comparison with the inheritance of physical characters. J. Anthrop. Institute 33, Pp. 204.

—'10. Nature and nurture. The problem of the future. Eugen. Lab. Lect. Ser. 6.

—'24. The life, letters and labours of Francis Galton. (2 vols.) London: Cambridge Univ. Press. Pp. 246 and 424.

Peters, W. '25. Die Vererbung geistiger Eigenschaften und die Psychische Konstitution. Jena, Pp. 368.

Peterson, Jos. '23a. The comparative abilities of white and Negro children. Comp. Psychol. Monog., 1, No. 5, Pp. 141.

—'23b. The use of a common unit in the measurement of race differences. Psychol. Bull. 20, 424–425.

—'28. Methods of investigating comparative abilities in races. Ann. Am. Acad. Pol. & Soc. Sci., 140, 178–185.

— and Barlow, M. C. '28. The effects of practice on individual differences. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. II, 211–230.

— and Lanier, L. '29. Studies in the comparative abilities of whites and Negroes. Ment. Meas. Monog. 5, Pp. 156.

—, — and Walker, H. M. '25. Comparisons of white and Negro children in certain ingenuity and speed tests. J. Comp. Psychol., 5, 271–284.

Phillips, B. A. '14. The Binet tests applied to colored children. Psychol. Clin. 8, 190–196.

Pintner, R. '23. Comparison of American and foreign children on intelligence tests. J. Educ. Psychol., 14, 292–295.

—'26. An empirical view of intelligence. J. Educ. Psychol., 17, 608–616.

—'29. The individual in school. General ability. In the Foundations of experimental psychology, 661–705. Worcester, Mass.: Clark Univ. Press.

— and Keller, R. '22. Intelligence tests of foreign children. J. Educ. Psychol., 13, 214–222.

Popenoe, P. '29. The child's heredity. Baltimore: Williams and Wilkins, Pp. 316. 417 titles in Bibliog.

Porteus, S. D. '17. Mental tests of delinquents and Australian aborigine children. Psychol. Rev., 24, 32–41.

— and Babcock, M. E. '26. Temperament and race. Boston: Badger, Pp. 364.

Poyer, G. '21. Les problems generaux de l'herédite psychologique. Paris, Pp. 302.

Pressey, S. L. and Cole, L. '19. Are the present psychological scales reliable for the examination of adults? J. Abnorm. Psychol., 13, 314–323.

— and Shively, I. M. '19. A practical information test for use with delinquents and illiterate adults. J. Appl. Psychol., 3, 374–380.

Pyle, W. H. '15. Mentality of the Negroes compared with whites. Psychol. Bull., 12, 71.

—'16. A manual for the mental and physical development of school children. Bull. Univ. of Miss., 17, No. 24. Pp. 32. (Revised 1920.)

—'18. A study of the mental and physical characteristics of the Chinese. Sch. and Soc., 8, 264–269.

Ranke, K. E. '06. Anthropologische Beobachtungen aus Zentralbrasilien. Aus den Abhandlungen der K. Bayer. Akademie der Wiss. II. kl. XXIV. Bd. I. Abt.

Reinhardt, J. M. '27. The Negro: Is he a biological inferior? Amer. J. Soc., 23, 248–261.

Reuter, E. B. '17–'18. The superiority of the mulatto. Amer. J. Soc., 23, 83–106.

Rice, C. H. '29. A Hindustani Binet—Performance Point Scale. Princeton, N. J.: Princeton Univ. Press. Pp. 196.

Rowe, E. C. '14. Five hundred and forty seven white, and two hundred and sixty eight Indian children tested by the Binet Simon tests. Ped. Sem., 21, 454–468.

Ruch, G. M. '25. The influence of the factor of intelligence on the form of the learning curve. Psychol. Monog., 34, Pp. 64.

—and Koerth, W. '23. Power vs. speed in Army Alpha. J. Educ. Psychol., 14, 193–208.

Schuster, E and Elderton, E. '07. The inheritance of ability. Eugenics Lab. Mems. I.

Scott, A. W. '30. A comparative study of responses of children of different nationalities and environments on intelligence and achievement tests. T. C. Bur. Publ. Pp. 30.

Shen, E. '25. The intellectual resemblance of twins. Sch. & Soc., '21, 601–602.

Shimberg, M. '29. An investigation into the validity of norms with special reference to urban and rural groups. Arch. of Psychol., 104. Pp. 84.

Slocombe, C. S. '26. The influence of practice in mental tests. Forum of Educ., 4, 173–179.

Snedden, D. S. '27. A study in disguised intelligence tests. (Interview form.) New York, T. C. Bureau of Publ., Pp. 48.

Spearman, C. '14–'15. The heredity of abilities. Eug. Rev., 6, 219-237.

Squires, P. C. '26. A universal scale of individual performance tests. Examination manual. Princeton, N. J.: Princeton Univ. Press, Pp. 198.

Starch, D. '20. Educational psychology. New York: MacMillan. Pp. 473.

Stockton, J. L. '22. Definition of intelligence in relation to modern methods of mental measurement. Psychol. Monog., 30, Pp. 118.

Stoneman, E. T. '29. State psychological clinic. Annual report for the year ending 30th of June, 1929. Perth Western Australia Dept. of Public Health, Pp. 26.

Symonds, P. M. '24. The intelligence of Chinese in Hawaii. Sch. & Soc., 19, 442.

Tallman, G. G. '28. A comparative study of identical and non-identical twins with respect to intelligence resemblances. 27th Yrbk. Nat'l Soc. Stud. Educ. Pt. I. 83–88.

Terman, L. M. '16. The measurement of intelligence. Boston, New York: Houghton Mifflin. Pp. 362.

—'25. Genetic studies of genius. (Vol. I. Pp. 648.) Mental and physical traits of a thousand gifted children. (Vol. II. Pp. 842—By C. M. Cox Miles, '26.) Early mental traits of three hundred geniuses. Stanford University, Cal.: Stanford Univ. Press.

—'28. The influence of nature and nurture upon intelligence scores: an evaluation of the evidence in Part I. of the 27th Yrbk. J. Educ. Psychol. 19, 362–373.

Terry, R. J. '29. The American Negro. Science, 69, 337–441.

Thomas, W. I. '12. Race psychology, standpoint and questionnaire. Amer. J. Soc., 17, 725–775.

Thomson, G. H. '21. The Northumberland mental tests. Brit. J. Psychol., 12, 201–222.

Thorndike, E. L. '05. Measurements of twins. Columbia Univ. Contrib. to Phil. and Psychol., 13, Pp. 64.

—E. L. '21. Intelligence and its measurement. J. Educ. Psychol., 12, 124–127.

—et al '26. The measurement of intelligence. New York, Columbia Univ. Publ. Pp. 616.

—et al '28. The resemblance of siblings in intelligence. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I., 41-53.

Vabalas–Gudaitis, J. '29. Ein einheitliches stummes Test system zur Bewertung von naturlichen Leistungen nach dem chronometrischen Prinzip. (A unified non-verbal system of tests for the measurement of natural capacities according to the chronometric principle.) Commentationes ordinis philologorum Universitatis Lituanae 5, Pp. 47.

Visher, S. S. '25. A study of the type of the place of birth and occupation of fathers of subjects of sketches in "Who's who in America." Amer. J. Soc., 30, 551–557.

Waardenburg, P. T. '29. Geestelijke eigenschappen bij tweelingen. (Character traits in twins.) Mensch en Maatschappij, 5, 17–34.

Walcott, G. D. '20. The intelligence of Chinese students. Sch. & Soc., 11, 474–480.

Wallis, W. D. '26. Race and culture. Sci. Mo., 23, 313–321.

Walters, F. C. '24. Language handicap and the Stanford revision of the Binet Simon tests. J. Ed. Psychol. 15, 276–284.

Walters, F. C. '27. Psychological Tests in Porto Rico. Sch. & Soc., 25, 231–233.

Wang, S. L. '26. Demonstration of the language difficulty involved in comparing racial groups by means of verbal intelligence tests. J. Appl. Psychol., 19, 102–106.

Waugh, K. L. '20. A comparison of Oriental and American students. Proc. Amer. Psychol. Assoc.

Whipple, G. M. '28. Editorial impression of the contribution to knowledge of the 27th Yrbk. J. Ed. Psychol., 19, 389–396.

Willey, M. M. and Herskovits, M. J. '27. Psychology and culture. Psychol. Bull., 24, 253–283.

Willoughby, R. R. '28. Family similarities in mental test ability. 27th Yrbk. Nat'l. Soc. Stud. Educ. Pt. I., 55–59.

Wilson, S. A. K. and Wolfsohn, N. K. '29. Organic nervous disease in identical twins. Arch. Neur. & Psychiat., 21, 477–490.

Wingfield, A. H. '28. Twins and orphans; the inheritance of intelligence. London and Toronto: Dent. Pp. 127.
Winship, A. E. '00. Jukes—Edwards: A study in education and heredity. Harrisburg, Pa.: Meyers, Pp. 88.
Wissler, C. '23. Man and culture. New York: Crowell Co. Pp. 371.
Wood, M. M. '29. Mental test findings with Armenian, Turkish, Greek and Bulgarian subjects. J. Appl. Psychol., 13, 266–273.
Woodrow, H. '19. Brightness and dullness in children. Phil.: Lippincott. Pp. 322.
Woods, F. A. '06. Mental and moral heredity in royalty. New York: H. Holt & Co. Pp. 312.
—'13. Heredity and the Hall of Fame. Pop. Sci. Monthly, 82, 445–452.
Woodworth, R. S. '10. Racial differences in mental traits. Science, 31, 171–186.
—'29. Psychology. New York: H. Holt and Co. Pp. 590. Rev. ed. (1st publ. 1921.)
Yeung, K. T. '21. Intelligence of Chinese children in San Francisco. J. Appl. Psychol., 5, 267–274.
Young, P. C. '29. Intelligence and suggestibility in whites and Negroes. J. Comp. psychol., 9, 339–359.

PART THREE

TEST SURVEY

PART THREE

Introduction.

In Part Two we indicated the difficulties attending studies which compare groups differing in any fundamental way. Chief among these were the factors of varying languages and varying cultures. It is our aim in Part Three to offer a compilation of tests which are relatively free from one of these obstacles, namely, language. The elimination of the 'local culture' handicap, although discussed earlier, is beyond the province of this survey.

We have attempted to assemble data on almost two hundred non-verbal tests covering the points which should prove useful to workers interested in adapting these tests for special purposes. For the sake of convenience the tests are grouped into blocks because it seemed that the items falling within each block were more closely inter-related than they would have been if classified in any other way. This is an arbitrary arrangement and admittedly cannot take all overlapping material into account.

The information on each test is often limited, due to the meagreness of the data obtainable. We have, however, been reasonably successful in recording the following: title of test, author, date, publisher or manufacturer, price, description of test material and problem. We have also stated whether the test was suited for individual or group testing or both, and whether the directions and responses were verbal or non-verbal. In the case of verbal directions, an opinion is given (usually our own) on the possibility of adapting them to pantomime.[66] No test is included, except those in group I, pages 78 to 85, which obviously could not be so adapted. A brief statement on administration—time and scoring —is given for each test, enough to indicate the feasibility of using the test for any particular purpose. A fuller notation is entered on standardization—the very basis of the mental test. The kinds of norms offered, the groups on which these norms were obtained, and the names of workers responsible for the standardizations are noted. In the case of several different standardizations, the most recent or the most extensive is reported. Data as to reliability and validity are also given wherever possible. Under the heading "discussion" are set down a few comments on each test gleaned from original books and articles. This might be considered "representative opinion" on each test, and is offered in lieu of our inability to state "what a test tests".[67] Each account closes with one or more references to articles bearing on the test. Pressure of time and space prevented these references from being exhaustive, but we trust they are sufficient to give the reader a point of departure in running down more details in regard to any particular test. If, in this compilation, we had tried to give information comparable to that found in Whipple's Manual, many volumes would have been filled.[68]

Published test information should be kept within the bounds of the profession, or else offered in such form that the unenlightened layman or that menace to high professional standards, the "six-weeks mental tester" shall not be able to use the psychologists' tools. The information included here is so general that the above risk will not be run.

The main introduction to this book outlined the range of this test survey. Lest the reader consider the choice of tests thus indicated too arbitrary, a fuller explanation may be given here. As they exist at present, both group and individual tests differ in the degree and manner in which they are free from verbal elements, so that three types may be identified:

	Directions	Responses[69]
1.	Pantomime	Non-verbal or manipulative
2.	Verbal	Non-verbal or manipulative
3.	Verbal	Verbal.

The first group comprises the tests aimed at, but the second was found to contain valuable material in which the language element might be omitted or adapted to pantomime. The third group is entirely omitted except in the case of a few verbal tests in which the testing principle gives promise of ready adaptation to concrete material.

The search for true non-verbal tests might be pushed into four definite fields: (1) emotional measurement; (2) pre-school and primary tests; (3) sensori-motor tests; and (4) special ability tests. The measurement of emotion, compared with intelligence testing, is still in its infancy; and while many non-verbal devices could be assembled, the results so far obtained are too uncertain to warrant report.[70] Pressure of time and space unfortunately prevented our examining primary tests, and this explains why so many group non-verbal tests are omitted,[71] as well as such individual tests with non-verbal elements as the various revisions of the Binet and the Detroit Kindergarten Test. We have included certain "low-level" tests, however, such as the Wallin Pegboards, but have omitted those which have been especially developed with the growth of the pre-school movement.[72] The third and fourth kinds of tests (sensori-motor and special ability) are, on the other hand, very decidedly drawn upon. Psycho-physical tests and even experimental techniques are included, although many are not standardized for intelligence testing purposes. If tests can be classified as dealing with the efficiency of (1) the sensory apparatus; (2) the motor apparatus and; (3) the brain mechanism lying between (1) and (2), then the sensori-motor tests in this survey lie on the two border lines of (3). Thus purely sensory tests such as visual acuity and pressure discrimination are omitted, as are purely motor tests of muscular strength. The border line includes tests of "coordination", "perception", etc.— such as the Tapping and the Manual Dexterity Tests.[73]

To those who question the wisdom of including tests of special ability such as the Stenquist test of mechanical ability and Seashore's Measures of Musical Talent we would say that here again it was hard to set a boundary. In the earlier stages of intelligence testing an attempt was made by use of the Stenquist to estimate the intelligence of men unable to handle linguistic material. Some of these special ability tests are not unlike the problem boxes, the puzzles[74] or the mechanical mazes, and have been included in many intelligence "scales." Furthermore, it could be argued that since this survey is but the first step toward covering the total field of mental measurement with the ultimate end of obtaining tests of "innate ability," the inclusion of such a test as the Seashore is highly appropriate, since it comes nearer to being a native ability test than any other. Special ability tests are designed for groups with definite criteria of success and this makes it likely that they will test differences in specific native capacities (training being allowed for) more accurately than tests of "general intelligence."

Finally, we can say that this "annotated list" of tests includes non-language materials developed since around 1911, the time when psychologists turned their attention from sensori-motor devices to measurements of the more complex thing called "intelligence." Some of the once acceptable intelligence tests are not included, because their authors insisted on their inadequacy.[75] Indeed, many of the tests reported are no longer valid or even extant, and are of interest only in historical perspective either as warnings or as "stepping stones to higher things." Elimination of all these now discarded tests would have distorted the picture of test development which can be traced in this survey. Indeed any list of tests brought up to date at one time will shortly have little but historical value for it will be rendered obsolete by the new tests constantly appearing on the market. If this survey succeeds in convincing the workers who produce these new tests that it is essential to give full and standard information on each test, it will have done much[76]. It is hoped that it will be useful in clinics, laboratories, and classrooms, as a short cut to source materials. If this proves to be the case, we can be forgiven for sacrificing detail of treatment to range of selection.

BIBLIOGRAPHY

References cited in the Introduction to Part III

Boody, B. M. '26. Testing immigrant children. Baltimore: Williams and Wilkins. Pp. 163.

Garrett, H. E. and Lemmon V. W. '24. An analysis of several well known tests. J. Appl. Psychol., 8, 424–438.

Otis, A. A scale for rating tests. Test Service Bulletin No. 13. Yonkers: World Book Co.

Pintner, R. and Paterson D. G. '17. A scale of performance tests. New York: Appleton. Pp. 218.

Ruch, G. M. '25. Minimum essentials in reporting data on standard educational tests. J. Educ. Res. 12, 349–358.

Squires, P. C. '26. A universal scale of individual performance tests. Princeton: Princeton University Press. Pp. 198.

Smith, H. L. and Wright W. W. '28. Tests and measurements. New York: Silver Burdett and Co. Pp. 556.

Part Three, Chapter I

VERBAL AND NON-VERBAL TESTS

By this time, the testing movement is so well established that intelligence is no longer "rated" or "estimated"; today it is measured by statistically calibrated tests and scales, which have a scientific basis. Although it is not our purpose to discuss the value of mental tests, a few statements as to their general advantages and limitations are in order, especially as regards the relation of verbal to non-verbal tests.

While the conditions of measurement in a classroom or clinic cannot be compared with the minutely controlled conditions in, let us say, a physics laboratory, nevertheless, psychology has acquired assurance on two points: (1) that intelligent behavior is measureable and (2) that, notwithstanding their shortcomings, some mental tests do serve adequately as its measuring instruments. Furthermore, the psychologist is his own severest critic, always insisting that interpretations of mental test results can be made only after all possible influencing factors have been accounted for. These factors include such things as the history, present emotional state, schooling, and social and economic conditions of the subject. Besides these, the training of the examiner, and the reliability and validity of the test must be carefully checked. (See next chapter.)

The content of tests necessarily varies with different testing programs. We have seen in Part Two that tests making use of experiences common to western culture, such as language, recognition of arithmetical symbols, coinage, household equipment, and so forth, can be used only within our civilization. If a comparison is wanted between a typical American group and one with a different cultural background, the test materials cannot be so localized. But there are many groups in our own civilization who do not speak our language, either because of the physical handicap of deafness, or because of a different national origin. If these are to be compared with English-speaking Americans they must be tested with non-verbal material. This stirs up the question: "Can the non-verbal test be used as a substitute for the verbal?" Will the mental operations called into play by the performance type of material be as symptomatic of intelligence as those stimulated by words, or are non-verbal tests measuring a special ability?

The psychologist accepts the principle, that, other things being equal, the more mental operations called into play,—according to Galton, the more "shafts sunk into the mind"—the more effective the testing scale. The performance test would seem to be a useful supplement to the verbal when the outward expression, the responses, are different. However, there is not always a wide gap between the responses: "performance" may extend all the way from gross manipulation, like throwing blocks, to subtle gesture, such as pointing or pushing with the finger to indicate a choice, as in Peterson's "Related Problems Test." This last kind of response however, involves behind it so high a degree of verbal organization and symbolism before the "principle" can be arrived at, that one hesitates to call it a non-verbal performance.

Directions of performance tests must also be taken into consideration: verbal instructions may form part of the test proper, as Pintner ('24) points out in the case of Primary group tests; or else they may be so familiar to the subject that to follow them is no task. Similarly, pantomime instructions may be so difficult that they are a problem in themselves; or else they may be expressed by such simple signals, that in the words of Squires ('26) they "put themselves across."

It might be possible to get expert judges to rate all of our tests for their degree of "verbality" and "non-verbality" in some such way as Herring ('21 and '24) did for his two scales (see page 68). Until this is done, we must be content with a rough two-fold classification in which (1) the verbal outclasses the non-verbal, and (2) the non-verbal outclasses the verbal. Perhaps the best way to distinguish between the two types is to compare extreme cases. Ideational or linguistic tests involve the use of complex symbolisms and can test a wide range of intelligence through words and abstractions. Performance or non-verbal tests on the other hand, since they demand only manual expression, are generally held to test only the lower levels of intelligence. As we have shown in the introduction (page 62), the true non-verbal test has pantomime instructions, but many tests with verbal directions are also called performance tests.

We shall now return to the main point at issue: Can the non-verbal test be used instead of the verbal, and if so, will the results of both tests be comparable? In the next chapter we outline a procedure for evaluating tests. We shall say here, as there, that judgments of tests must be based on statistical research and not on mere opinions. However, the views of eminent psychologists, as well as experimental evidence, will be given below for they are bound to be illuminating on the matter of verbal versus non-verbal tests. This question is not a new one, and is so closely bound up with its correlative problem, the nature of intelligence, that statements of opinion, of necessity, touch upon both problems at the same time.

OPINIONS

1. TERMAN.

We have seen (page 13) that Terman does not believe in different *kinds* of intelligence (such as ability to handle tools as against ability to solve mathematical problems), but insists on different *levels* of intelligence ranging from the very simple to the very complicated. In 1921 he indicated emphatically that performance tests do not measure intelligence as he defines it, since "an individual is intelligent in proportion as he is able to carry on abstract thinking." The most successful intelligence tests would have to be those which involve the use of language and other symbols, and therefore non-verbal tests cannot reach the higher intellectual levels. In proportion as the latter have something in common with the best verbal tests such as arithmetical reasoning, completions, opposites, analogies, etc., are they successful as tests of intelligence.

2. THORNDIKE.

Thorndike ('21) offers a twofold point of view about test responses: they can be considered as (a) ability to handle a situation which has all its elements present; and (b) ability to handle a situation in abstraction. There is no sharp dividing line between these two kinds of ability. Equally true is it to say that the correlation between them cannot be perfect, because no two kinds of mental activity ever correlate perfectly. Thorndike also emphasizes the *complexity* of intelligence, which necessitates a wide assortment of tests.

3. PINTNER.

Pintner ('21), too, stresses the point about the complexity of intelligence: "Intelligence is shown in dealing with things as well as with symbols, in handling a paint brush or chisel as well as a pencil." One kind of test will therefore not do to measure intelligence; many reactions are needed. (See also page 12.)

4. F. L. WELLS.

F. L. Wells ('27) thinks that performance tests do not measure the same thing as verbal but that the wide range of abilities which represent a person's intelligence requires an *ensemble* of tests for accurate measurement. (This opinion is also shared by Squires '26.)

5. DEARBORN, SHAW AND LINCOLN.

Dearborn, Shaw and Lincoln ('23) point out that there is a spread of abilities within the performance field itself. Even a mechanical ability test may serve as an instrument to measure abstract thinking. In other words, mental operations cannot be deduced from the materials employed to test them. Then too, because simple demands are usually made by performance tests, one should not think that it is impossible to set up a performance situation calling for analytic and synthetic thinking. So far, performance tests fail to discriminate above the age of ten or thereabouts.

6. BRIGHAM.

Brigham[11] thinks that performance tests may tap a particular ability, the training for which might be more specialized than that for verbal tests. If this were so, performance tests could not be considered as "intelligence tests."

EXPERIMENTAL EVIDENCE.

1. ARMY DATA.

The first intimation that verbal and non-verbal tests do not measure the same thing came from the work of the army psychologists (Yerkes' '21). The non-verbal group test (Beta) correlated with the Stanford-Binet .74 (p. 387) and with the verbal group test (Alpha) .806. The five most significant intelligence tests in the Army performance scale correlated with the Stanford-Binet .84. The mechanical ability test (Stenquist) yielded very low correlations with the Stanford-Binet and with the Alpha: indeed, feeble-minded men often rated high on the Stenquist. (Note: While a mechanical ability test may tap abstract intelligence, as stated above, it is generally considered as a test of a "special ability", apart from intelligence.)

2. Pintner Non-language.

The Pintner Non-language Test has been found to correlate .78 with Liu's composite criterion on 235 children. Pintner ('24) concluded, after using it on another group, that similar abilities were not measured by verbal and non-verbal tests.

3. Thorndike Non-verbal.

The coefficient of correlation between Thorndike's Non-verbal Test and the Army Alpha was .685. Thorndike ('19) writes that there is a possibility that different aspects of intelligence are being measured by the two kinds of tests.

4. Pintner-Paterson Performance Scale.

When Pintner and Paterson ('17) set out to devise a performance scale which would meet the needs of those people to whom verbal tests could not be given, they stated that they were guided by three principles: (1) the complex character of intelligence, which made many tests necessary. (See above.) (2) Stern's definition of intelligence, as "quick adaptability to solving new problems". (3) Each test should require no language response. Pintner and Paterson do not report coefficients of correlation with the Stanford-Binet, or other standard criteria of intelligence, but other workers, Johnson and Schreifer, ('25) obtained a coefficient of .83 for children under ten years of age.

5. Liu Non-verbal.

Liu ('22) appears to have been the first to make an exhaustive examination and analysis of group non-verbal tests. After setting up an almost unassailable standard of requirements for his battery, he proceeded to validate each element in it against a criterion composite equally demanding. His requirements were fourteen-fold: The tests should involve no language responses; their materials should exclude school training as much as possible; they should call for no special abilities, but should appeal to all types. Tests should be indicative of growth from year to year; they should be valid, reliable, and supplied with simple administrative instructions which could be handled by non-specialists. Scoring should be objective, materials inexpensive and easy to handle; alternative forms should be provided. But most important for our discussion, he required that the tests should measure "a large number of unlike or differentiating traits". The ideal way would be to "measure every trait that contributes to intelligence and to give each trait a weighting proportional to the contribution to the total intelligence".

The criterion against which Liu validated his tests was a weighted composite of age, teachers' estimates, school marks, school progress and test scores. By correlations, partial and regression equations, the test elements in the scales were checked against this criterion and the foremost valid tests selected. His final coefficient of correlation with this criterion was .88. (high)

6. Herring verbal non-verbal scale.

Herring ('21 and '24) reported on the distribution of (1) intelligence elements from concrete to abstract; and (2) test elements from non-verbal to very verbal.

He selected test elements from the standard intelligence scales and arranged them into two series: (1) a concrete extreme to an abstract extreme, and (2) a non-verbal extreme to a highly verbal extreme (on the basis of expert judgments). He correlated the validity of the series against an elaborate composite criterion of intelligence, consisting of: Stanford-Binet, educational age, relations between chronological age and grade reached, National Intelligence Test, teachers' estimates, Thorndike Alpha and Kelley-Trabue Completion Alpha. His conclusions are: 1. Manifestions of intelligence cannot be divided into two groups, concrete and abstract. 2. The range of responses from most concrete to most abstract is a *continuous* one and not a disparate one. 3. In general, verbal tests call for abstract thinking, non-verbal tests for concrete action. 4. For this reason verbal tests are superior to non-verbal in diagnostic power at the upper levels.

Again, the fundamental weakness which can be charged against the performance test is that it has failed to develop into difficulty levels of the sort which would call for complicated abstract thinking.

7. Squires' Scale of Individual Performance Tests.

In 1925, Squires set himself to investigate just this problem. He devised a scale of fourteen tests, tried them out on an inferior group of feeble-minded males, and a superior group of college juniors and seniors. As a result of his research, he writes:

> "Evidence growing out of the present investigation shows that it is quite possible to obtain high correlations between verbal and non-verbal tests. The historical fact of the low correlation between the two types of tests we believe to be explained not necessarily by assuming that verbal and non-verbal tests measure different aspects of intelligence, but by taking into consideration the possibility that inadequate technique in the construction and selection of non-verbal tests may have been the chief factor responsible for the absence of a high correlation between the two types."

Here, then, is evidently the pioneer effort to correct the chief ailment of the performance test. If the Squires scale, or some other upper range performance series, could be improved from the standpoint of reliability and validity, obviously the performance test could, and would, come into its own, almost, if not altogether on a par with the verbal test.

SUMMARY.:

A survey of opinions and researches suggests the following needs:

1. More tests should be built with non-verbal directions, and more of the existing so-called non-verbal tests should be standardized in pantomime situations.

2. Devices such as the profile technique should be developed for recording the qualitative aspects of performance on non-verbal tests.

3. Complete information on standardization should always be available, preferably supplied with the test material.

4. Interpretations must not be based on the results of a single test, nor on too narrow a range of tests. Different kinds of abilities must be sampled by the different tests in the performance scale. (See Appendix I, No. 67 on the determination of independent abilities as tested by different types of material.)

5. Provision must be made for testing the upper levels of intelligence with performance tests.

6. The discriminative value of each test in a scale should be statistically determined. (See Arthur '25.)

If these six points could be satisfactorily attended to, many of the objections which have grown up against performance material might prove to be unfounded. In other words if non-verbal tests could conform to the requirements of a good test, there is every reason to assume that they would be good measures of intelligence. For example, Liu's non-verbal scale which was based on fourteen rigid specifications, succeeded in agreeing closely with an unassailable criterion of intelligence. (See also Drever and Collins '28.)

CONCLUSION:

Adverse opinion of performance tests can be traced not to the fact of their being non-verbal, but to certain inherent weaknesses which would disqualify any test as a measure of intelligence. As put by Dodd ('26): "By increasing both the amount and complexity of the mental functioning sampled by non-verbal tests, they may approach the efficiency of verbal tests with respect to validity and reliability." Arthur ('25) says: "It would seem that performance scales when constructed according to Binet principles, can be trusted to do the work of a Binet scale in measuring general intelligence."

Even at this stage of development, non-verbal tests are useful supplements to verbal tests. Certain inherent advantages lie with performance tests: (1) They usually appeal to the subject. (2) When individual, they provide the examiner with opportunity to observe in the subject such factors as "persistence," "ability to vary and adapt methods to changing situations," "emotional reactions in the face of apparently baffling situations," etc. (3) Performance tests tap certain abilities which are not tapped by verbal tests.

BIBLIOGRAPHY

References cited in Part III, Chapter I.
Verbal and Non-Verbal Tests.

Arthur, M. G. '25. A new point performance scale. J. Appl. Psychol. 9, 390–416.

Dearborn, W. F., Shaw, E. A., Lincoln, E. '23. A series of formboards and performance tests of intelligence. Harvard Monog. Educ. 1, No. 4. Pp. 64.

Dodd, S. C. '26. International group mental tests. Princeton, N. J.: Princeton Univ. Store, Agents. Pp. 101.

Drever, J. and Collins, M. '28. Performance tests of intelligence. Edinburgh: Olive & Bays. Pp. 52.

Johnson, B. J. and Schriefer, L. '22. A comparison of mental age scores obtained by performance tests of the Stanford Binet. J. Educ. Psychol. 13, 408–417.

Liu, H. C. '22. Non-verbal intelligence tests for use in China. T. C. Contrib. to Educ. Pp. 126.

Herring, J. P. '21. Verbal and abstract elements in intelligence examinations. J. Educ. Psychol. 12, 511–517.

—'24. Herring revision of the Binet Simon tests and verbal and abstract elements in intelligence examinations. Yonkers, N. Y.: World Book Co. Pp. 72.

Pintner, R. '21. Intelligence and its measurement. A Symposium. J. Educ. Psychol. 12, 139–143.

—'24. Results obtained with the non-language group tests. J. Educ. Psychol., 15, 473–483.

— and Paterson, D. G. '17. A scale of performance tests. New York: Appleton. Pp. 218.

Squires, P. C. '26. A universal scale of individual performance tests. Princeton: Princeton University Press. Pp. 198.

Terman, L. M. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol., 12, 127–133.

Thorndike, E. L. '19. A standardized group examination of intelligence independent of language. J. Appl. Psychol. 3, 13–32.

—'21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 12, 124–127.

Yerkes, R. M. (Ed.) '21. Psychological examining in the United States Army. Mem. Nat'l. Acad. Sci., 15, Pp. 890.

Wells, F. L. '27. Mental tests in clinical practice. Yonkers, N. Y.: World Book Co. Pp. 315.

Part Three, Chapter II

STATISTICAL BASES OF TEST INTERPRETATION

While recognizing that every human being is a unit unto himself, and that consequently intelligence testing cannot be a "strictly quantitative science" (F. S. Freeman '29), the fact remains that before a test can be useful for individual diagnosis, it must be "standardized"—that is, the degree of success of given groups on the test must be ascertained. The performance of each future individual can then be compared with the average and the distribution of scores on the same test of the group to which he belongs. It is for this reason that any critical judgment of a test must be based on group statistics. The discussion below will outline the most important factors to be considered when knowledge about the value of any test is wanted. It will also show what complexities await the investigator who wishes to make his data comparable with that of others.

As has already been admitted in the Introduction, some of the tests included in this survey are but poorly adapted for the purpose of testing intelligence. The question arises: Is it possible to turn to authoritative sources and find decisions "ready made" as to the relative virtues of each test? Apparently not, for as Ruch ('28) says:

> "Even if the cards were all on the table, the selection of the "best" test would present grave difficulties. These decisions must be made and are being made daily, but it is an open question whether any living being possesses the exact knowledge required to make such decisions in anything approaching a scientific manner."

and according to Kelley ('27, p. 30):

> "Though many would be inclined to accept the judgment of some eminent psychologist that the test was valid in preference to the figures of an uncertain tabulator and interpreter of correlation data, nevertheless, it is not too much to demand that the validity of forthcoming tests be adequately supported by indubitable correlation results, as well as by the opinions of their authors."

SAMPLING

The research worker can draw no conclusions as to the value of a test unless he knows on what "sampling" it has been tried out, because different groups may perform very differently on the same test. (For a definition of "samling" see Appendix I, No. 64). The question of sampling will crop up in the discussion of each statistical procedure below, so the following bare generalizations will suffice here: A sampling, to be representative of the population, must be *large;* it must be *random,* and its *range,* as well as its average, must be known. No factors must be allowed to operate within a sampling that would cause it to be "selected" i. e., to be other than it is assumed to be. (See Burks and Kelley '28, p. 9 and page 37 of this volume.)

VALIDITY

Before a research worker can tell whether a test is a "test of intelligence" he must know about its diagnostic power—i. e., its *validity*. The theory underlying the general intelligence test of today is that mental capacity is indicated by ability to profit from everyday experiences. But, as Dunlap ('30) says: "We can find many performances that children on the average "pick up" but do they show ability to "pick up" other things? To be standardized a test must be given to subjects whose mental ability is known, having already been gauged by some acceptable "measure", such as school marks or respected judgments. This previous measure is known as the "criterion," and the validity of the new test will depend upon the extent to which its findings agree or "correlate" with the rankings made by the criterion. In general the same people who rank highest and lowest in one test will rank highest and lowest in the other, and the ranking of those in between will remain relatively unchanged. The closer the agreement of rankings on the two tests, the higher the coefficient of correlation or "r" and hence the greater the validity of the new test. (For definition of "r" see Garrett '26, p. 144.)

There are apt to be statistical snags here, however. As Hull ('27) points out, the size of an "r" depends on the formula by which it was obtained. Different formulae are available, some being better adapted to one set of figures, some to another. The most easily applied formula is not always the most suitable. Then too, a coefficient of correlation can never be rightly interpreted, unless it is accompanied by other data such as age, range and variability of the group on whom it was obtained. Herring ('24) writes:

> "Increase in the S. D.'s[77] involved in a correlation, results in an increase of correlation which is paralleled by no fact, difference, or change in human nature; it is a function of increase in the reliability of the group, and with regard to the correlation existing between two traits it is wholly spurious. An "r" of .4 in a group having a standard variability of five months of M (criterion composite) may mean just the same degree of mutual implications as an "r" of .8 in a group having a much higher standard variability."

To gauge the validity of a test, then—i. e. to learn how well its results agree with the findings of an accepted criterion—one must know not only what formula was used to obtain the correlation coefficient, but also certain facts about the group on which the test was used; the number of the cases, their age and character, and their "spread" or variability.[77]

RELIABILITY

Having answered the question as to how well a certain test measures what it sets out to measure, the research worker must ask: Will the test be as good next week as it is today? In short, how reliable is it? A test may be ineffective on second testing because of some inherent weakness in itself: it may be too short, too vague in its directions or scoring rules so that different examiners get different results with it. People, too, have inherent weaknesses: they become excited, nervous, timid, mentally "blocked". In other words, unreliability may depend either on a lack of objectivity in the test itself, or else on the daily variations of the subjects. The first and second results from a "good" test will tend to agree more closely than will those from a poor test. This agreement is discovered by using the correlation technique, the size of the coefficient indicating how much a test can be counted on to behave the same way twice.

Again, as in the case of validity, there should be presented certain other aspects of the data by which the reliability coefficient was arrived at. Different formulae are available for different sets of data, and the significance of the coefficient depends not only on the formula used, but on the type of test offered, the variability of the group, and so forth. Kelley ('21) indicates the need for cautious inspection of the reliability coefficient:

> "To secure a reliability coefficient of .40 from a group composed of children in a single grade is probably indicative of a greater, not less, reliability than to secure a reliability coefficient of .90 from a group composed of children from the second to the twelfth grades."

To increase the reliability of a test, certain factors must be properly dealt with, such as objectivity of scoring, directions, length, and the like. Toops ('26) points out that a thirty minute test may be better than nothing at all, but is much worse than a three hour test; and that to tap a subject's real range of ability, a wide sampling of tasks must be offered.

One other limitation of the reliability coefficient must be born in mind. When properly interpreted, it will predict the behavior of the test from group to group and from time to time, but it will not indicate the error of the test on an individual score. The amount of this error can be predicted by the P. E. (Probable Error) of the estimate of an individual score. This P. E. is derived through the use of a statistical formula which takes into account the S. D. (Standard Deviation) of the distribution as well as the reliability coefficient. (See Garrett '26, p. 184f). It indicates how much the error of the second measurement would be if the test were repeated, decreasing when the S. D. decreases, and when the coefficient of reliability increases. Kelley cautions us never to lose sight of

> ". . . . the ubiquitous probable error. A tin can on a dog's tail is a very effective reminder to the dog that he is not a free agent, and a probable error attached to an intellect in such a manner that it proclaims itself whenever the imagination runs rampant, would be equally serviceable . . . ." ('27, p. 19)

NORMS

If a test is both valid and reliable, two of the most important requirements in its standardization will have been met. The research worker will then inquire into the actual distribution of scores obtained by using the test on a particular sampling. Qualitatively the sampling must include various types in the same proportions as they exist in the whole community. (Pintner and Paterson '17) and quantitatively, it must include a large enough number of cases to present a "normal curve of distribution". (For definition of the normal curve, see Garrett '26, p. 74.) This tabulation results in a series of tables or "norms" for various categories within the sample such as age, grade, sex, and so forth.

Norms have been reported chiefly in two ways. The first was to set down merely what the average performance of a group would be without considering the standing of the best or the poorest members. Later interest shifted from group performance to the individual variation within the group. This gave rise to the second method of presenting norms, namely recording how much the individual excelled or fell below the average of the group. The first is exemplified by the age scale, in which, for instance, a task, which is passed by about three quarters of the seven-year-olds, is taken as the norm for that year. Values in terms of points are then assigned to different tasks and these converted into "age norms." A mental age or M. A. of seven is "seven-year mentality"; which means the performance which would be executed by a normal group of seven year-olds on an assortment of tasks such as those in the Binet Scale. Some seven-year-olds do better, some worse than the average seven-year-old. The ratio between a child's actual age, and his mental age is his intelligence quotient, or I. Q. An average child of any age gets an I. Q. of 100. If a seven-year-old performs as an average eight-year-old would, his I. Q. will be greater than 100, and *vice versa,* if he performs as an average six-year-old would, his I. Q. will be less than 100.

The second outstanding method of reporting norms deals in terms of the variation of the group. It is another method of expressing how much the individual differs from the average. This procedure is illustrated by the "percentile method" and is the preferred procedure for use with adults because with them mental growth can no longer be used as a basis for scoring. The scores of individuals in any group are ranked from highest to lowest and the total number is then divided into a hundred equal parts. A person in the lowest hundredth or "first percentile"[78] will have 99 per cent of his group above him; the one in the fiftieth percentile would have as many above him as below; and the one in the 98th percentile will have only 2 per cent above him. The value of having a percentile ranking in addition to some other indication of norms lies in the possibility of combining scores on one test with scores on another test which may be expressed in different terms. John Jones, for instance, earns a credit of 6 minutes on a problem box, a score of 20 correct moves on a certain formboard, 80 points on a group test, and a mental age of 10 years on the Stanford-Binet. How are these results, all expressed in different measures, to be dealt with? Percentile rankings offer one solution: If the examiner knows the Subject's percentile ranking on one test (with respect to his group), he can compare this with the Subject's ranking on a second, or any number of tests.

COMBINING SUB-TEST SCORES

In regard to the matter of combining scores on different kinds of test materials, warnings have been sounded recently by such authorities as Kelley and Brigham. These writers point out the fallacy of adding up credits for the several parts of a test. Kelley ('27, P. 122) has shown that certain traits are relatively independent of each other, such as verbal, quantitative and spatial intelligence.[79] Brigham ('30) comments:

> "The findings of Kelley as to the probable existence of independent traits rather effectively challenge current test procedure, as most tests combine verbal and mathematical materials indiscriminately. As most of the non-verbal tests deal with spatial relations, one should not attempt to find similarities between scores in verbal and non-verbal tests . . . ."
>
> "It is apparent that scores in tests found to be independent should not be added."

SUMMARY

Before he can label any test "efficient", the research worker must observe the following points: 1. The test must have been standardized on a representative sampling and full details given on the standardization. 2. These should include the number and description of cases, age, range and variability of the groups from which data on reliability, validity and norms were obtained. 3. The coefficients indicating validity, (diagnostic value) and reliability (trustworthiness) must be reported, together with the formulae by which they were obtained, and also their probable errors or "P. E's." 4. The various sub-tests must have been separately validated, so that non-equivalent scores will not have been added to make a simple total.

If the information on any test conforms to the above requirements, it is possible to tell on what kind of subjects it can be used. The research worker who wishes to measure a single case can then choose a test which has been standardized for individual testing, and can feel confident that it is efficient. (N. B. Standardization for individual testing involves stricter requirements than for group testing). If he wishes to compare groups, he must make certain that he tests a sufficient number of individuals in each group to form a "normal curve of distribution," after which he may choose his test according to the above criteria and rest assured that it is comparatively sound.

BIBLIOGRAPHY

References cited in Part III, Chapter II
Statistical Bases of Test Interpretation.

Brigham, C. C. '30. Intelligence tests of immigrant groups. Psychol. Rev. 37, 158–165.

Burks, B. S. and Kelley, T. L. '28. Statistical hazards in nature-nurture investigations. 27th Yrbk. Nat'l Soc. Stud. Educ., Pt. I. 8–38.

Dunlap, K. '30. Mental tests. Progr. Educ. 7, 57–67.

Freeman, F. S. '29. Intelligence tests and the nature-nurture controversy. Sch. & Soc., 31, 830–836.

Garrett, H. E. '26. Statistics in psychology and education. New York: Longmans, Green and Co. Pp. 317.

Herring, J. P. '24. Herring revision of the Binet-Simon tests and verbal and abstract element in intelligence examinations. Yonkers, N. Y.: World Book Co.

Hull, C. L. '27. The correlation coefficient and its prognostic significance. J. Educ. Res. 15, 327–338.

Kelley, T. L. '21. The reliability of test scores. J. Educ. Res. 3.

—'27. The interpretation of educational measurements. Yonkers: N. Y. World Book Co. Pp. 363.

McCall, W. A. '22. How to measure in education. New York: Macmillan. Pp. 416.

Pintner, R. and Paterson, D. G. '17. A scale of performance tests. New York: Appleton. Pp. 218.

Ruch, G. M. '25. Minimum essentials in reporting data on standard educational tests. J. Educ. Res. 12, 349–358.

Squires, P. C. '26. A universal scale of individual performance tests. Princeton: Princeton University Press. Pp. 198.

Terman, L. M. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 12, 127–233.

Toops, H. A. '26. Theoretical principles underlying tests and measurements in vocational education. Proc. Nat'l. Educ. Ass., 837–889.

Yule, G. U. '21. Critical notice of Brown and Thomson's "Essentials of Mental Measurement." Br. J. Psychol. 12, 105–106.

ANNOTATED SURVEY OF TESTS

Learning Tests which Require Language.

Group	I.	Learning Tests, Nonsense Syllables, etc.

Tests Relatively Independent of Language.

Group	II.	Memory and Learning Tests.
Group	III.	Sensory Tests.
Group	IV.	Sorting Tests.
Group	V.	Special Ability Tests and Mechanical Puzzles.
Group	VI.	Constructive Ability Tests.
Group	VII.	Picture Tests.
Group	VIII.	Construction Tests.
Group	IX.	Formboards.
Group	X.	Mazes.
Group	XI.	Tests Involving Recognition of Form, Color, Weight.
Group	XIII.	Spatial Relations.
Group	XII.	Group Tests.

I

LEARNING TESTS WHICH REQUIRE LANGUAGE

No. 1 Crossline Test A

Devised by D. P. Macmillan.
Date: 1911
Published by Stoelting.
Price: Ten cents.
Material: Paper on which a large X is drawn, with figures inserted between the four angles of the X.
Problem: To associate the figures, or letters, with position and form.
Individual test; but could be adapted to group testing.
Directions: Verbal; pantomime possibly.
Responses: Non-verbal, but calls for use of figures.
Time: No time limit.
Standardization: Brigham and Morganthau found this test to have little discriminative value for normal children. (Morganthau '22, p. 30)
Discussion: A modification of this test appears in the Goddard and Terman revisions of the Binet-Simon scale. Crossline A and B served as a basis for the "code test" in year XVI of the Stanford revision.
References:
Healy and Fernald '11
Goddard '11
Schmitt '15
Terman '16
Dearborn and Brewer '18
Morganthau '22
Bronner, Healy, Lowe, Shimberg '27

No. 2 Crossline Test B

Devised by W. Healy and G. M. Fernald.
Date: 1911
Published by Stoelting.
Price: Ten cents.
Material: Paper on which is drawn a four-lined figure. Between these lines numbers are inserted.
Individual test; four trials; could be adapted to group testing.
Problem: To associate numbers with position and form in the design.
Directions: Verbal; pantomime by adaption.
Responses: Non-verbal, but call for use of number.
Time: No time limit
Scoring: 1. Trial at which all spaces are numbered correctly, or 2. Sum of correct figures on each trial, plus 9 for each trial not given (Bronner Healy *et al* '27)

Standardization: This test was standardized by Hazeltine, Lowe and Shimberg on 1225 cases. The three scores are distributed among three age groups only, ranging from nine to seventeen years. (Bronner, Healy, Lowe and Shimberg '27, p. 34)

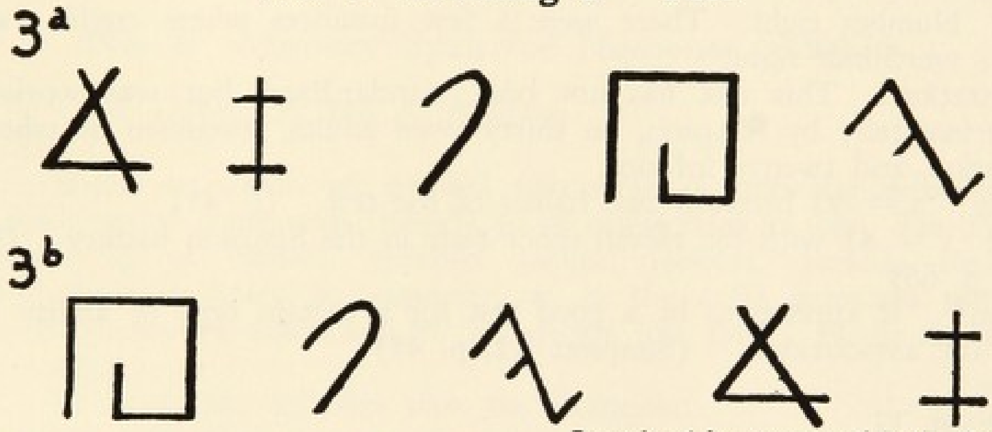
Validity: Brigham obtained a "diagnostic value" on this test by subtracting the percentage successful among defectives from the percentage successful among normals. Only two tests in the Binet-Simon series exceeded this test in value. (Brigham '17, p. 185)

Discussion: This test is a modification of Crossline Test A. Bronner, Healy, *et al*, have found but little discriminative value for both this test and Crossline A. ('27, p. 179) Yerkes and Foster also offer a modification of this test.

References:

Healy and Fernald '11
Schmitt '15
Terman '16
Brigham '17
Dearborn and Brewer '18
Morganthau '22
Yerkes and Foster '23
Bronner, Healy, Lowe and Shimberg '27

No. 3 Learning B Test



Reproduced by courtesy of Little, Brown & Co.

Devised by W. Healy and A. F. Bronner.
Date: 1927
Published by Stoelting.
Price: Twenty cents.

Material: Paper on which six rows of five symbols are arranged in varying order; Key of names for the experimenter.

Problem: To associate symbol and name and reproduce orally.

Individual test: Three trials. This test is adapted to group testing, but would involve written language responses.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal.

Scoring: Total number correct on three trials (or allowed for).

Standardization: This test was standardized by Regensburg, on 633 cases, and reported in the Bronner-Healy Manual. There is an absence of age to age norms, and the authors comment that the test is too short to be reliable.

The association between symbol and sound seems to call for a special ability, as there is a wide scattering of scores at each age. ('27, p. 199)

Discussion: "Success in this particular type of rote learning is a special ability often in distinct contrast to the subject's powers with ideational material." (Bronner, Healy, *et al*, '27, p. 199)

References:

Bronner, Healy, Lowe, Shimberg '27

No. 4 Learning Pairs Test

Devised by E. L. Thorndike (?).

Date: 1912

Material: Four lists of characters or symbols, each associated with a word; ten pairs to each list.

Problem: After being given opportunity to study the association of word with symbol to write down corresponding word for each symbol when seen alone.

Individual or group test.

Directions: Verbal; pantomime adaptation difficult.

Responses: Verbal.

Time: One minute to study connections for first two trials; one and a half minutes for last two trials. (p. 9 f)

Scoring: Number right. There were a few instances where credit was given for a word half right.

Standardization: This test has not been standardized, but was worked with experimentally by Simpson, on thirty-seven adults, seventeen of whom were superior, and twenty inferior.

Reliability: r = .93 between two halves of the test. (p. 45)

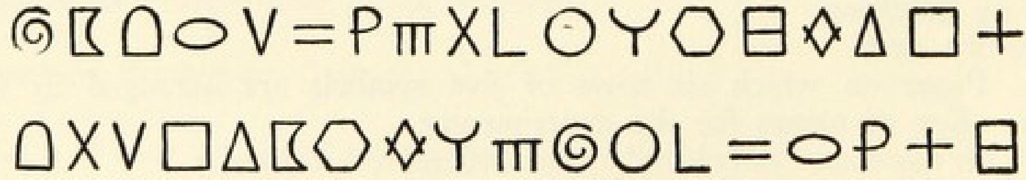
Validity: r = .41 with the eleven other tests in the Simpson battery. (Simpson '12, p. 66)

Discussion: "It appears to be a good test for a certain type of ability to form and use associations." (Simpson '12, p. 45)

Reference:

Simpson '12

No. 5 Learning Z Test



Reproduced by courtesy of Little, Brown & Co.

Devised by W. Healy and A. F. Bronner. (This test is a modification of their Learning A Test.)

Date: 1911

Published by Stoelting.

Price: $1.70 per hundred.

Material: Paper on which are six rows of symbols, eighteen to a row; key for eighteen symbols.

Problem: To associate symbols with digits, visually presented.
Individual test; three trials. This test could be adapted to group testing.
Directions: Verbal; pantomime with difficulty, as they involve use of numbers.
Responses: Verbal; non-verbal by adaptation.
Scoring: Number right on all three trials.
Standardization: This test was standardized by Shimberg, Lowe and Meehan on 772, 856 and 772 cases, for age, grade and in percentiles. Range: from eight to eighteen years, Grades: three to twelve. Age to age norms are absent, a wide scattering being reported at each age. (Bronner, Healy, Lowe and Shimberg '27, p. 80)
Discussion: The learning test was originally offered to gauge "the powers of attention and the ability of the subject to establish a comparatively easy set of associations." (Healy and Fernald '11). Bronner, Healy, *et al,* report that they found no correlation between this test and school retardation, although the learning processes involved in these arbitrary associations are apparently identical with some of the processes necessary in learning to read. ('27, p. 197)
References:
Bronner, Healy, Lowe, Shimberg, '27
Morganthau '22

No. 6 Memory Span for Nonsense Syllables

Devised by C. B. Davenport and A. Anastasi.
Date: 1929
Material: Sheets on which are printed two series of nonsense syllables, which are made up of consonants thought to be universally in use. The first series is made up of random syllables (sokpe, lolsotka, pumno, etc.). The second series is based on sequence, as in the cube imitation test, rather than on qualitative differences of the sounds (do, do da, do de da, do de da di . . . . etc.)
Problem: To repeat the syllables after the Examiner.
Individual test; not adapted to group testing.
Directions: Pantomime.
Responses: Verbal, but not linguistic.
Time: Unlimited.
Scoring: The number of lines in every respect correctly reproduced: no partial credits.
Standardization: This test has not been standardized.
Discussion: "The second test is better than the first as a universal test, in that in many languages all syllables end with a consonant. Also the test is one of recalling sequence and pattern, rather than qualitative sound differences, that are introduced by a variety of consonants . . . .
The chief psychological processes involved in performance in this test are: Attention, memory, perseveration, adaptability, analysis and synthesis." (Davenport)
Reference:
Davenport, unpublished report.

No. 7 Memory Material

Devised by E. M. Achilles.
Date: 1920
Handled by Stoelting.
Price: $0.80 per hundred for each set.
Material: Two forms of fifty nonsense syllables in each form, consisting of three-letter words: two consonants with a vowel between, e. g., zof, dej, zaf, etc.
Problem: To recall or to recognize in the second list, words which were previously presented in the first list.
Individual or group test.
Directions: Verbal.
Responses: Verbal.
Time: Fifty seconds for study series.
Scoring: For recall: total correct. For recognition: total number—twice number of errors—omissions, if any. (For fuller discussion see Achilles '20, p. 26 f.)
Standardization: Ninety-six adults and some 600 children in grades IV through VIII were tested by Achilles. Age and grade norms, together with average deviations and mean variations, are given. (Achilles '20) The material, however, has not been standardized as an intelligence test, although progressive age-grade increases are revealed.
Discussion: Achilles was interested in studying the functions of recall and recognition in children and adults. These nonsense syllables form a part of the memory material which she used in her comparative study. She concludes that recall and recognition must be thought of as one continuous process rather than two. The essential difference is that the threshold for recognition is lower than for recall. The reader is cautioned to bear in mind which aspect of memory he is trying to test.
References:
Achilles '20

No. 8 Serial Learning Test

Devised by W. H. Pyle.
Date: 1923
Handled by Stoelting.
Price: $5.00
Material: Five series of ten nonsense syllables in each. (For use with a Jastrow tachistoscope).
Problem: After a one-second exposure for each ten syllables, the Subject is required to write down in their right order as many as he can recall. Five exposures are allowed for each series.
Individual or group experiment.
Directions: Verbal.
Responses: Verbal; or non-verbal.
Time: Enough, but no more than is required to write down all the words *readily* recalled.

Scoring: Efficiency score = sum of separate scores for the five series.
Standardization: This material is reported in Pyle's Laboratory Manual on learning, and is not offered as a test of intelligence. It is included here with other learning tests because of its promise of ready adaptability to different test situations. (Pyle '23, p. 121 ff.)
Reference:
Pyle '23

No. 9 Rational Learning Test

Devised by Joseph Peterson
Date: 1916
Published by author
Material: Written instructions. No printed test blanks are necessary.
Problem: To complete learning (of association of number and letter) as soon as possible, depending upon discovery of an underlying principle.
Individual test.: Nine items for adults, fewer when used for children. Not adapted to group testing.
Directions: Verbal, but may be given in the Subject's own language.
Responses: Verbal.
Time: No time limit but the time required is noted.
Scoring: Time, repetitions, errors, (logical, perseverative, unclassified) total.
Standardization: Norms are expressed in percentiles for repetitions, time and errors. 113 college students were studied by Peterson. Peterson and Harrelson published norms for eight, nine and ten-year-olds and Peterson and Lanier for twelve-year-olds. Results have been published on the application of this test to 570 white and 698 Negro children.
Reliability: $r = .67 \pm .05$ (Peterson and Lanier '29)
Validity: $r = .37$ with Binet,
$r = .53$ with Myers Mental Measure. (Peterson and Lanier '29)
Discussion: Peterson ('18) claims that this test measures accurately and easily all one's relevant responses to a definite, objective situation. He was interested to see the relation between rational learning and trial and error learning as well as intelligence. Reliability and validity are relatively high as compared with other learning tests. (Peterson, Lanier and Walker '25) This test and the Mental Maze and Disc Transfer tests are presumably independent of school training: Each test stresses factors that have little in common with the others. They do not measure a general ingenuity factor. (Letter)
References:
Peterson '18, '20 a and b, '22, '23, and '28, and letter to writers
Peterson, Lanier and Walker '25
Peterson and Lanier '29

No. 10 The Disc Transfer Test

Devised by Joseph Peterson
Material: Rectangular piece of paneling 8 x 6 inches upon which are painted three circles, 1, 2, 3; each 2½ inches in diameter; five wooden discs

decreasing successively in diameter from a little less than 9 inches to about ¾ of an inch.

Problem: Fore exercise: The three largest discs are stacked in circle 1; the Subject is told to get them over to circle 3, moving only one at a time and never placing any disc upon a smaller one, and leaving in the same order. The real test follows the same procedure in more complicated form.

Individual test; not adapted to group testing.

Directions: Verbal.

Responses: Non-verbal

Scoring: Time and number of moves.

Reliability: Few data.

Validity: Few data. See Peterson and Lanier '29, pp. 82-83.

Standardization: This test was tried on 119 whites and 92 Negroes in Nashville; 17 whites and 40 Negroes in Chicago. Norms are expressed in percentile scores for time, repetitions, and errors.

References:

Peterson, Lanier and Walker '25

Peterson and Lanier '29

No. 11 The Mental Maze Test

Devised by Joseph Peterson

Material: Fore-exercise card on which are drawn mazes having numbers at the various intersections. Directions.

Problem: Fore-exercise: to "thread" the maze in view orally by choosing which of two numbers, called by the Examiner, indicates the right intersection so that finally the maze is threaded without error.

Real Test: To do the same thing with words instead of numbers only without any "maze" in view.

Individual test; not adapted to group testing.

Directions: Verbal but may be given in Subject's own language.

Responses: Verbal.

Scoring: Time, repetitions and errors.

Validity: See Peterson and Lanier '29, pp. 82-83.

Reliability: See Peterson and Lanier '29, pp. 82-83.

Standardization: This test was tried out on two groups of whites (N = 119 and 17) and on two groups of Negroes (N = 192 and 40). Percentiles for time and number of moves are given.

References:

Peterson, Lanier and Walker '25

Peterson and Lanier '29

No. 12 Related Problems Test (Bead Test)

Devised by John C. Peterson; modified by J. L. Graham.

Date: 1920

Handled by authors

Materials: A string of beads mounted on a steel wire from which E and S take alternate draws.

Problem: To discover the principles which will produce a desired result in the final draw.

Individual test; a series of problems.

Directions: Verbal

Responses: Manipulative; but principle must be thought out in words.

Time: No time limit

Scoring: Trials and time.

Standardization: Peterson experimented with this test on 46 college students. Graham developed it in his studies on the comparative abilities of whites and Negroes.

Validity: Graham reports moderate but positive correlations with semester's scholarship combined with Otis self-administering test scores.

Discussion: Graham says "Such tests have decided value for inter-racial testing and may be an improvement upon tests in which mere insight or recognition of relational ideas is required. As to their universal serviceableness, I would call your attention to the great difficulty of abstract and symbolic thinking. Tests can be construed upon my principle which will be easier than these. The principle, I believe, holds possibilities for race testing."

"The purpose is to discover new relationships. The principle differs from most of the current intelligence tests in that it is not a test of the recognition of relational ideas based upon excellency of the organism and previous experience, but in that it aims to start with examples in which the solution is obvious, immediate and clear, i. e., in which the control is perceptual and to proceed to related examples of gradually increasing difficulty and complexity in which the solution is beyond the student's perceptual range, and in which a response to relational ideas is a necessary pre-requisite of a prompt and accurate solution of the related problems. It is therefore a learning test, but at a level where the responses require abstract, symbolic and relational reactions." (letter to writers)

References:

John Peterson '20
Graham '30
Graham, letter to writers

II

MEMORY AND LEARNING TESTS

No. 13 Ideational Learning Board (Demonstration Board)

Devised by P. R. Farnsworth and W. H. Roberts.
Date: 1926
Handled by Stoelting.
Price: $16.50
Material: Double peg board with twelve holes bored on both sides, resting on a support and equipped with forty-five colored T-shaped pegs which can be inserted and removed at will.
Problem: Subject tries to ascertain an essential element in a total situation previously arranged by Examiner. Problems of indefinite complexity can be built up.
Individual or group experiment.
Directions: Verbal; pantomime doubtful.
Responses: Verbal; non-verbal possibly.
Standardization: This board with its wide range of problems awaits standardization.
Discussion: Farnsworth and Roberts, in offering this board, believe that problems for different ages can be developed and standardized. They think that this board will prove useful to test inductive reasoning.
Reference:
Farnsworth and Roberts '26.

No. 14 Peg Design Test

Devised by New York Bureau of Analysis and Investigation.
Date: 1917.
Handled by toy stores, and Stoelting.
Price: $1.25; also a modification at $.50.
Material: Box with checkerboard top in which there are thirty-three holes for the insertion of pegs.
Problem: To reproduce a design so that pegs may be "jumped off the board."
Individual test: two parts. This test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: About ten minutes.
Scoring: Part I: Time in seconds, for total effort; number of times design is reproduced.
Part II: Time for arranging and jumping pegs, and number of attempts to arrange pegs.

Standardization: This test was standardized by the New York Bureau of Analysis and Investigation on 317 cases. Norms are expressed for time, and per cent success, in terms of age. Range: from seven to fourteen years.

Discussion: "It has been found that for younger children, it is chiefly a test of rote memory, but older children grasp the idea and reason it out. Besides learning powers it also, in some degree, tests remote memory." (Bronner, Healy, Lowe, Shimberg '27, p. 233)

References:

New York Bureau of Analysis and Investigation '17
Carlisle '18
Bronner, Healy, Lowe and Shimberg '27

No. 15 Association Test (Pictorial)

Devised by P. C. Squires.

Date: 1926

Produced in the Psychological Laboratory, Princeton University.

Material: Three pairs of pictures glued to one-inch cubes, one pair to a cube. Demonstration board, 9 x 2 x 1/4 inches, on which are three cubes mounted on pivots, etc. (For details, see Squires '26, p. 28 ff.)

Problem: After a learning period in which Subject can learn three pairs of pictures, one member of each pair being on the front face and the other member being on the back face of a cube, the task is to determine what pictures are on back faces of *N* cubes.

Individual test: Series A, three sets of six items; learning series only. Series B. three boards of six items each. This test is not adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: No time limit.

Scoring: Time; moves; successes.

Standardization: This test was reported by Squires on the basis of results with fifty Princeton men, and fifty feeble-minded men. Norms are not yet established. The test is considered applicable to low-grade and primitive people.

Validity: $r = .19$; for scoring (by successes) and —.28 (time); criterion, the higher score on Princeton or Thorndike intelligence examination; subjects, the fifty college men.

$r = .17$; (successes); and .12 (time); criterion is Stanford-Binet mental age for the fifty feeble-minded subjects. (p. 184)

Discussion: While Squires has made no attempt to conjecture what mental processes are involved in his various tests, he feels that this test might be worth investigating in regard to its worth at the lowest ages, where it might be valuable in differentiating individuals. He advises that more items be devised. (Squires '26, p. 193)

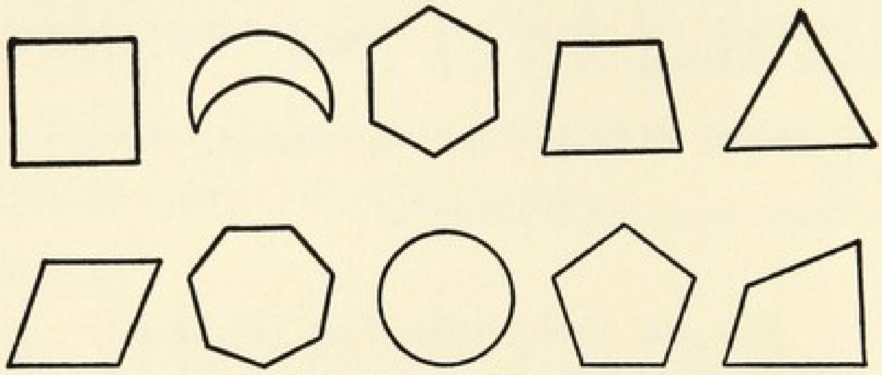
References:

Squires '26

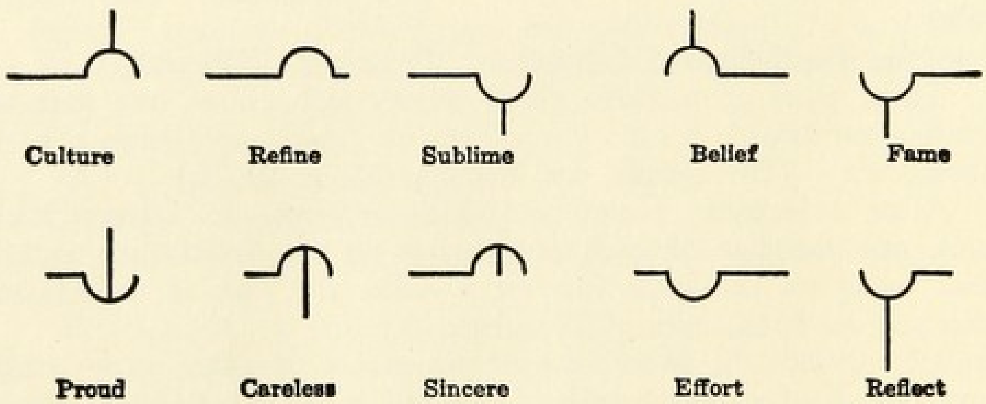
No. 16 Associative Learning

Devised by L. W. and F. L. Kline.
Date: 1927

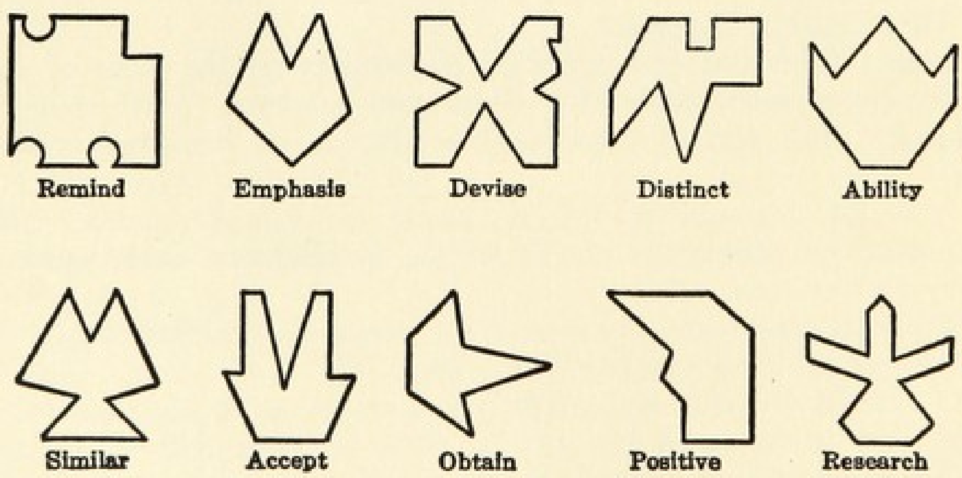
SERIES I



Set *a*. Meaningful figures



Set *b*. Less meaningful figures

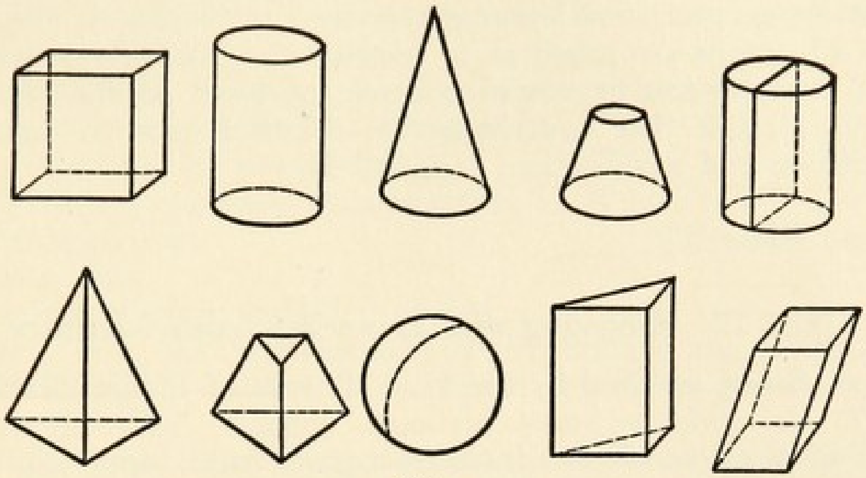


Set *c*. Nonsense figures

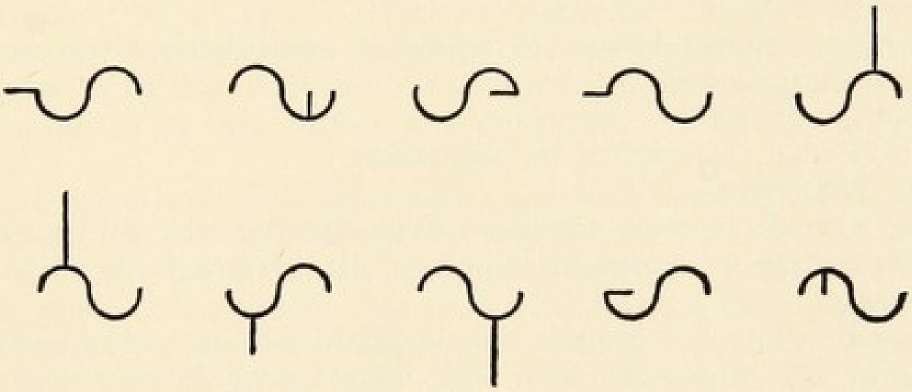
Material: A number of 6 x 8 inch cards, on each of which are drawn ten unusual outlines; cardboard 10 x 12 inches; stop watch.

Problem: To reproduce a design after a ten-seconds exposure. Individual or group experiment.

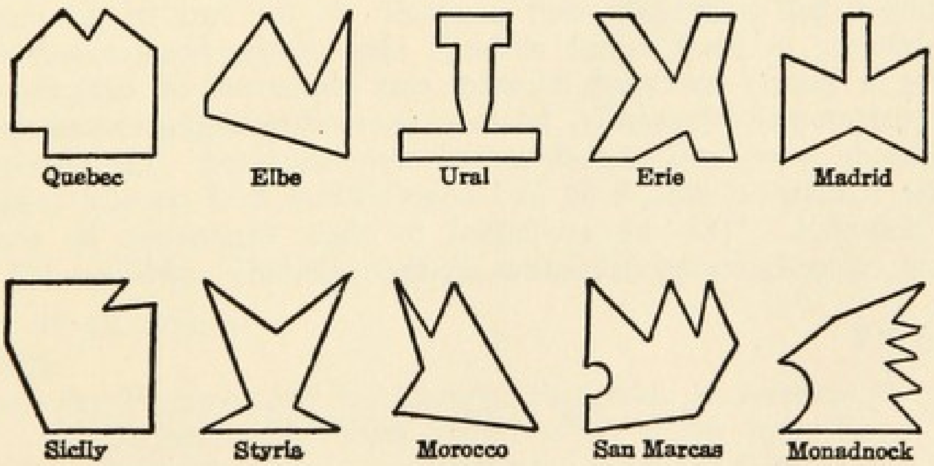
SERIES II



Set *a*



Set *b*



Set *c*

Reproduced by courtesy of Ginn & Co.

Directions: Verbal.
Responses: Verbal, or non-verbal by adaptation.
Standardization: This material has not been standardized for mental testing, but is included here because of its suggestiveness for learning experiments and development of new learning tests.
Discussion: Questions are raised as to whether the Subject remembers geometrical and other figures by name or shape, or by a combination of the two, or by other cues. The underlying idea of the exercise is capable of indefinite variation and modification. (p. 219 f)
Reference:
Kline and Kline '27

No. 17 Copying and Drawing from Memory

Devised by A. Binet; modified by the U. S. Bureau of Public Health.
Date: 1917
Material: Papers on which are drawn five geometrical forms: square, diamond, double trapezoid, bent wire figure, and allied-Ziehen figure, placed one at a time before Subject.
Problem: After copying figures, to reproduce immediately from memory. Two practice items allowed .
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Time: Five seconds exposure for easier drawings; ten seconds for more difficult.
Scoring: According to agreement of judges, each drawing was placed in one of three grades.
Standardization: This test has not been standardized, but was reported upon by Mullan, ('17) who gave it to 295 literate and illiterate aliens. He felt that it was dependent on previous training.
Discussion: Mullan felt that formal schooling was the greatest single determinant of success on this test. He urged the importance of finding out if the Subject had previously used a pencil. In the case of an immigrant, a poor drawing in itself meant nothing. But if the Subject had had school training, a poorly executed drawing was suggestive of defective "analytic and visual retentive powers". Mullan observed individual peculiarities in the way pencils were held and drawings were executed. He also noted that illiterate Ruthanian girls, who had never before held pencils, could produce good drawings. This he attributed to their experience in embroidering different colored geometrical figures on their clothing. (Mullan '17, p. 71 f.)
Reference:
Mullan '17

No. 18 Cube Imitation

Devised by H. A. Knox; modified by R. Pintner.
Date: 1914
Manufactured by Stoelting; also by Marietta.

Price: Knox's version \$4.50; Pintner's modification \$1.25 (Stoelting); \$3.00 (Marietta).
Material: Knox's set consisted of four colored one-inch cubes (red, blue, green and yellow) fastened to a board, and one black cube for tapping. The Pintner Paterson modification consists of five 3.5 cm. cubes of similar color and size, unattached.
Problem: To imitate the Examiner tapping in a certain order.
Individual test: Twelve different arrangements. This test is not adaptable in its concrete form to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: No time limit.
Scoring: Number of times passed or failed.
Standardization: This test was standardized by Pintner and Paterson ('17) on 900 school children. Norms are expressed for raw scores in terms of age, percentiles, quartiles. Range: from five years to adult.
Reliability: "The table shows a good distribution with relatively little scattering. .... The amount of variation of the middle fifty per cent is fairly small and fairly constant at all ages." (Pintner, Paterson, '17, p. 136)

"In our experience the variability of performance on repeated trials has led us to believe that the test is unreliable. Having formerly used it very extensively we now employ it very seldom." (Bronner, Healy, *et al*, '27 p. 196)

"It is significant for age levels only as a member of a considerable group of tests." (F. L. Wells '27, p. 138)

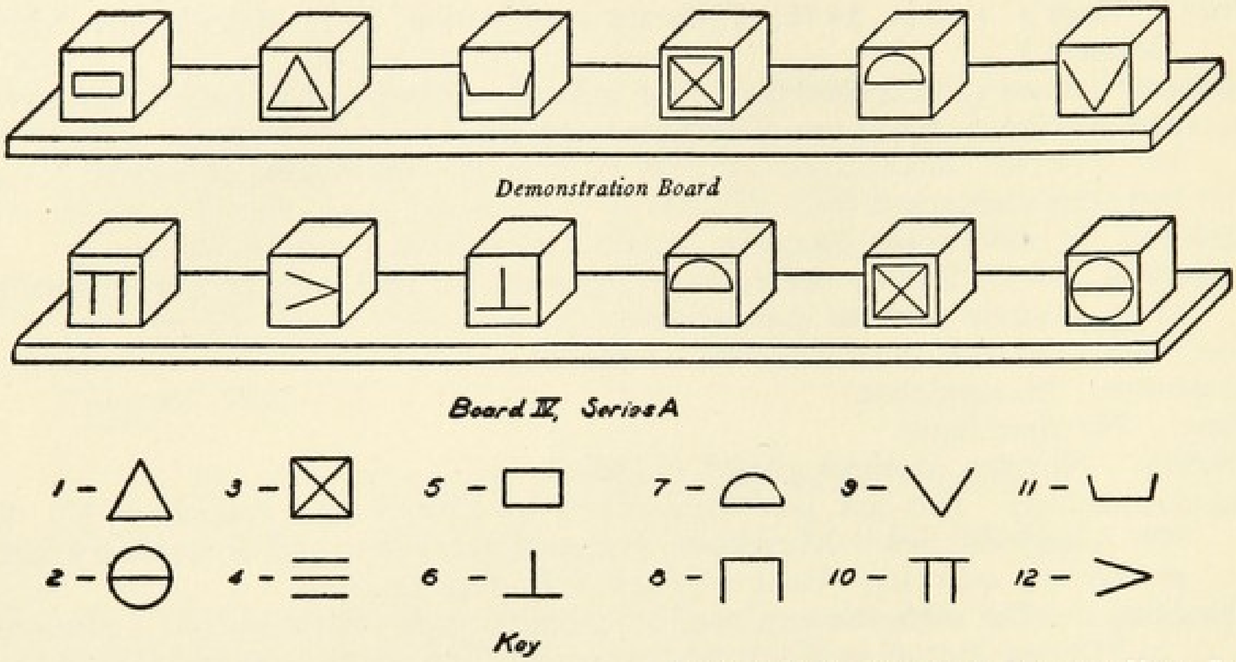
Discussion: "Much depends on the process of memorizing the patterns. Overt response gives relatively little clue as to what this has been. Visual, auditory and kinaesthetic processes readily enter. The test may be regarded as one of intelligence in so far as high scores reflect systematic use of such cues." (Wells '27, p. 138)

References:

Knox '14
Pintner '15
N. Y. Bureau of Analysis and Investigation '15
Pintner and Paterson '17
Rachofsky '18
Carlisle '18
Yoakum and Yerkes '20
Yerkes, (Ed) '21
Starr '23
Mateer '24
Arthur '25 and '28
Squires '26
F. L. Wells '27.

No. 19 Geometric Association Test

Devised by P. C. Squires.
Date: 1926.
Produced in the Psychological Laboratory, Princeton University.
Materials: Six demonstration cubes, mounted on pivots to a board, 18 x 2 x ¼ inches. Series A contains four boards. In Series B the cubes are not attached to any board. Counter box of twelve compartments. (For details, see Squires '26 p. 69 ff).



Reproduced by courtesy of Princeton Univ. Press

Problem: After a preliminary learning series, to associate N geometric designs, presented at a time. N is not less than two, or more than five.
Individual test: Series A, learning, three sets. Series B, test, four sets with increasing number of items. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative
Time: No time limit.
Scoring: Time; moves; successes.
Standardization: Squires experimented with this test upon fifty Princeton seniors, and fifty feeble-minded men in Vineland. Norms are not standardized.
Validity: r = .06 (successes); —.25 (time) for Princeton group, criterion being the higher score on either the Thorndike or Princeton intelligence examination.
r = .02 (successes); and — .02 (time) on Vineland group, criterion being Stanford-Binet mental ages. (p. 184)
Discussion: "The pictorial and geometric association tests appear to be extremely poor." (Squires '26, p. 193)
Reference:
Squires '26

No. 20 Imitation Test (Block Test)

Devised by H. A. Knox.
Date: 1914
Handled by Stoelting.
Price: $4.00
Material: Five three-inch wooden cubes.
Problem: To imitate models built by the Examiner.
Individual test; not readily adaptable to group testing.

Directions: Verbal; or pantomime by adaptation.
Responses: Manipulative.
Standardization: Knox placed this test at the six-year level, but made no definite standardization. No one, since his time, appears to have been interested in it as a test, and it still lacks standardization.
Discussion: This test is not to be confused with Knox's "Cube Imitation Test". In the latter the Subject imitates the Examiner's order of tapping cubes; in the "Imitation Test", the Subject imitates the structure built out of three cubes by the Examiner.
Reference:
Knox '14

No. 21 Learning Figures Test

Devised by C. E. Seashore; modified by C. B. Cornell.
Date: 1917
Material: Paper on which are odd-shaped designs adapted from the Seashore manual.
Problem: After ten seconds exposure, the Subject is asked to reproduce designs.
Individual test: Three trials. The test is adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Time: Ten seconds exposure.
Scoring: Credit for total number right on three trials; one-half point allowed for reversed drawings.
Standardization: This test was standardized on 550 public school children Norms are expressed in terms of age. Range: six to fourteen years.
References:
Seashore '08
C. B. Cornell '17

No. 22 Memory for Objects Test

Devised by F. W. Ellis.
Date: 1915
Manufactured by Stoelting.
Price: $1.20
Material: Ten small metal toy objects: shoe, dog, hat, etc.
Problem: To remember in order the names of objects on a tray, after the objects are removed
Individual test; group test by adaptation.
Directions: Verbal; pantomime by adaptation.
Responses: Verbal; non-verbal by adaptation.
Scoring: Record order named, allowing ten points for each object named. Penalize one point for each misplacement.
Standardization: This test was standardized by Dewey, Child and Ruml on 500 Jewish cases. Norms are expressed for raw scores in terms of age and sex. Range: from nine to thirteen years.

Discussion: These objects possess inherent attraction for children and hold the attention of low grade subjects. They also approximate concrete situations in everyday life. (Bronner, Healy, *et al,* '27, p. 201)

References:

Ellis and Bingham '15
Dewey, Child, Ruml, '20
N. Y. Bureau of Analysis and Investigation '17
Weidensall, '16
Bronner, Healy, Lowe and Shimberg, '27

No. 23 Memory of Objects Test

Devised by C. B. Cornell.
Date: 1917
Material: Fifteen familiar objects fastened to white cardboard, thirty-seven centimeters square: button, comb, crayon, spectacles, key, knife, marble, nail, nickel (coin), pen, pencil, scissors, screw, stamp.
Problem: After five seconds exposure, to name from memory the objects seen.
Individual test; adaptable to group testing with use of written word.
Directions: Verbal; pantomime by adaptation.
Responses: Verbal or non-verbal.
Scoring: Number remembered.
Standardization: This test was standardized by Cornell on 550 public school children. Range: six to fourteen years. Norms are expressed in terms of age.
Discussion: This is similar to Ellis' ('15) Memory for Objects Test, (No. 22) standardized by Dewey, Child, Ruml, ('20) and Mullan's "Visual Apprehensions Test" (No. 30) etc.

References:

Ellis and Bingham, '15
C. B. Cornell '17
Mullan '17
Dewey, Child, Ruml, '20

No. 24 Multiple Choice Method

Devised by R. M. Yerkes; adapted by Münsterberg for mental examining.
Date : 1916
Handled by Stoelting.
Price: $167.50
Material: The essential features of this apparatus are: A bank of twelve hardwood keys which can be operated by the Experimenter from one side of a screen and by the Subject from the opposite side; a signal or switchboard on which are mounted twelve miniature lamps, each of which is connected with one of the keys; also twelve jacks, each connected with one of the keys, and which may be connected with a buzzer; a wooden screen to separate Subject from Experimenter; a box on which the above parts are mounted and in which are placed the electrical mechanisms for the signal board.

As adapted for mental examining purposes, the material consists of:

a. (Münsterberg) small cards, on which are printed three lines of letters arranged in a certain relation.

b. (Yerkes-Rossy) circles and squares instead of letters. A given card bears four lines of symbols, in each of which appear circles and one square. The relation of the latter to the circles is the same for each of the four lines.

Problem: (Yerkes-Rossy) To discover and describe relations of varying degrees of difficulty.

Individual test: Four relational problems. Adapted in Yerkes-Rossy and Münsterberg forms for group testing.

Directions: Verbal; pantomime uncertain.

Responses: Manipulative, written or verbal.

Scoring: Average score is total number of reactions, or number of repetitions before solution is reached for problems one to four.

Standardization: Yerkes worked with this apparatus experimentally on forty-eight normal adults and thirteen defectives. Standardization is not reported. He says it is adapted to human, infra-human, mature, immature, normal, defective, or diseased subjects.

Discussion: "The solution gives opportunity for ideational reaction, though not necesarily depending upon it; ability to deduce from a series of experiences the principle which binds them together and explains them." (Yerkes '21, p. 372)

"Some such procedure as the relational method may profitably be adapted to the needs of the mental examiner as a means for measuring for practical purposes such ideational characteristics of human subjects as number of ideas, their quickness of development and value." (Yerkes '21, p, 386)

"Experience indicates that the relational test has considerable practical value in mental examining as well as varied values as a method of research." (Yerkes '21, p. 379)

"In every sense the method, in each of its forms, is a measure of intellectual function. I have used it for groups, by means of Yerkes-Rossy cards and projection on screen." (letter to the writers)

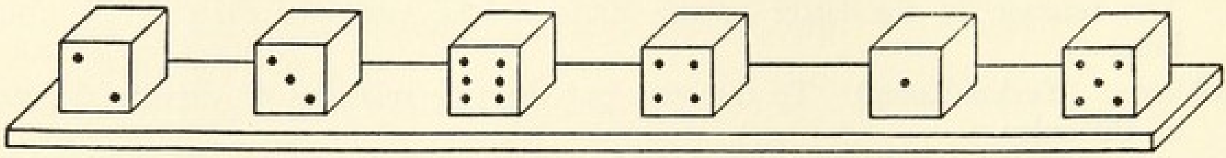
A variation of method was devised by Dashiell, who introduces a board in which are drilled twenty-five circular holes. Into these holes, wooden discs are to be placed which can be easily lifted out. The subject is instructed to search for a piece of ribbon hidden under different discs, all falling within a given geometric design. The subject is to try to discover the design. (Dashiell, '28, p. 516 f.)

References:

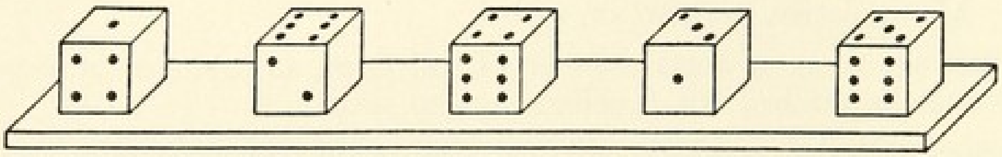
Yerkes and Rossy '17
Yerkes '21b, and letter to writers.
Bronner, Healy, Lowe & Shimberg '27
Dashiell '28.

No. 25 Quantitative Association Test

Devised by C. C. Brigham; modified by P. C. Squires.
Date: 1926.
Produced in the Psychological Laboratory, Princeton University.



Board I, Series A



Board I, Series B

Reproduced by courtesy of Princeton Univ. Press

Material: Six boards on which cubes are mounted. Demonstration board, mounted in the same manner; single demonstration cube. Counter box with eleven compartments. (For complete description, see Squires '26, p. 37 ff.)

Problem: After practice in learning three pairs of dice patterns, task consists in determining the sum on back and bottom faces of cube when only one side is shown.

Individual test; Series A, (learning) three boards, each containing six items. Series B, three boards, each containing five items. This test is not adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: No time limit.

Scoring: Time; moves; successes.

Standardization: This test was tried out by Squires on fifty Princeton seniors, and fifty feeble-minded males of Vineland Training School. Norms were not established, but the test seems to be applicable through feeble-minded and superior levels.

Validity: $r = .19$, (based on successes) and $-.50$ (time) for fifty Princeton men, criterion being the higher score on either the Thorndike or Princeton intelligence tests.

$r = .39$ (for successes) and .39 (time) for the feeble-minded men, criterion being Stanford-Binet mental age. (p. 184)

Discussion: The test ranked high for the feeble-minded group, both in respect to its multiple correlations and its regression score ratings. This test also appears to be rather effective for superior subjects." (Squires '26, p. 192)

Reference: Squires '26

No. 26 Series Forward

Devised by P. C. Squires; after H. A. Knox (in Cube Imitation).
Date: 1926
Produced in the Psychological Laboratory, Princeton University.
Materials: A four bar xylophone, 12 x 6 x (3½) 3 inches, with four colored bars; two wooden mallets, 7½ inches long. (Squires '26, p. 13 f)
Problem: To reproduce a series of strokes in the order presented by the Examiner.
Individual test: Three practice groups, with three to six series in each; eleven test groups, with five series to each group; each "span" increasing. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Successes until Subject fails in all five items of one group.
Standardization: Squires worked with this test experimentally on fifty Princeton students, and fifty feeble-minded males of Vineland. Norms are not yet established.
Validity: r = .26 (for successes) on Princeton group. Criterion was the higher score on either the Thorndike or the Princeton college entrance intelligence tests.
r = .19 for the feeble-minded men, criterion being the Stanford-Binet mental age. (p. 184)
Discussion: "It should be emphasized that no attempt has been made in the present study to conjecture what mental processes are involved in the performance of the various tests. . . .The Series Forward, Series Reversed, and Free Analogies tests do not contain enough intra-test variables to permit of including these tests in the partial correlation lists." (Squires '26, p. 188 f.)
Reference:
Squires '26.

No. 27 Series Reversed

Devised by P. C. Squires. (after Knox).
Produced in the Psychological Laboratory, Princeton University.
Date: 1926
Material: A four-bar xylophone (12 x 6 x (3½) 3 inches) with four colored bars; two wooden mallets, 7½ inches long. See Series Forward test, No. 26.
Problem: To beat out a series of strokes in reversed order to that tapped by the Examiner.
Individual test: Three practice groups containing three to five series in each; nine test groups each containing five series; "span" gradually increasing. This test is not adaptable to group testing.
Directions: Pantomime.

Responses: Manipulative.
Time: No time limit.
Scoring: Successes.
Standardization: Squires experimented with this test on fifty Princeton seniors and fifty Vineland feeble-minded males. Norms are not yet established.
Validity: (successes)
r = .23 for Princeton group. Criterion is the higher score on either the Thorndike or Princeton intelligence examination.
r = .04 for Vineland group. Criterion is Stanford-Binet mental age (p. 184)
Discussion: "The Series Reversed Test perhaps draws too freely upon special capacity or special ability to be as useful for general diagnostic purposes as the Series Forward Test. In respect to multiple correlation ratings, the Series Reversed Test appeared to better advantage in the college grade than in the feeble-minded population." Squires reports also that all of the Vineland and some of the Princeton men experienced difficulty with the pantomime instructions. (Squires '26, p. 191)
Reference:
Squires '26

No. 28 Spot Pattern Test

Devised by P. C. Squires. (after Whipple)
Produced in the Psychological Laboratory, Princeton University.
Date: 1926
Materials: Whipple portable tachistoscope and stand, pattern cards, a peg board, and spherical-headed pegs. (For details, see Squires, '26, p. 124 f.)
Problem: Given a spot pattern exposed for three seconds, and the same number of pegs as there are spots in the pattern, to reproduce the pattern.
Individual test: Series A, four practice items, five test items. Series B, five test items. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Time; moves; successes.
Standardization: Squires experimented with this test on fifty Princeton seniors and fifty feeble-minded males from Vineland. Norms are not established.
Validity: r = .25 (successes) and — .38 (time) for Princeton group, criterion being the higher score on either the Thorndike or Princeton intelligence test.
r = .11 (successes) and .07 (time) for the Vineland group, criterion being the Stanford-Binet mental age. (p. 184)
Discussion: Squires thinks that the chief task of this test is to determine pattern or configuration. He finds the peg board superior to pencil and paper responses, as the former sets its own problem and is readily understood by very young children and primitive peoples. Pencil and paper on the other hand is undesirable because it involves school factors. (Squires '26, p. 131 f.)

"The Spot Pattern Test should develop in time into a first rate test for all mental levels, ... should be useful in the investigation of racial differences." (Squires '26, p. 192.)

Maxfield writes that the tachistoscope in this test is very unsatisfactory. He thinks that a standard tachistoscope of some kind should be substituted. (letter)

References:

Squires '26

Maxfield, letter to writers

No. 29 Substitution Test

Devised by J. E. Lough; modified by a number of experimenters, notably Woodworth and Wells, Pintner, Woolley, Pyle, Dearborn and Starch.

Date: 1911

Handled by Stoelting.

Price: Per hundred: Whipple's $1.50; Pyle's $1.50; Woolley's $3.20; Woodworth-Wells $1.50; Link $1.50; Pintner-Paterson $0.80; Averill's $0.80.

Material: Printed blanks containing a varied arrangement of different geometrical forms; key strip, showing a digit or letter to be associated with each geometric form. Pyle used forty series of five place numbers as set down by Whipple; Woodworth-Wells used five geometrical figures, each repeated twenty times; Pintner used only the upper half of the Woodworth-Wells sheet; Woolley used four blank pages of nine different kinds of geometrical figures, fifty figures on a page; Link has a form combining mixed letters and numbers.

Problem: Sometimes it is associating the digits with the symbol and reproducing the former when the latter is exposed; sometimes *vice versa*. (Digit-symbol or symbol-digit test)

Individual or group test:

Directions: Verbal; pantomime by adaptation. In the latter case Pintner found it necessary to introduce a fore-practice period.

Response: Non-verbal.

Scoring: Time and accuracy. Pintner and Paterson penalized one-fiftieth of total time for each error.

Standardization: This test was tried on eleven adults by Woodworth-Wells; on 5202 children from fourteen to eighteen years by Woolley; on 666 school children ages four to sixteen by Pintner and Paterson; and on 3729 males and 4023 females, ages eight to eighteen by Pyle. Norms are expressed for raw scores, in terms of age and sex. Pintner and Paterson also give percentile and quartile distributions.

Reliability: Pintner and Paterson report that the distribution of scores is exceptionally good, showing very little scattering. (Pintner and Paterson '17, p. 133).

Discussion: Whipple gives three different forms of this test. It was developed primarily to study the learning process, or the ability to form arbitrary associations, and is akin to the Healy-Bronner learning tests. As the latter point out, however "the growth curve" as stated by Pyle (p. 47) "is practically a straight line for the girls up to age sixteen and boys up to seventeen." The Healy-Bronner learning tests, on the other hand, show practically uniform norms from grades three to seven after which there is a wide gap. These authors are unable to explain the cleavage between the lower and upper levels. (Bronner, Healy Lowe and Shimberg '27, p. 198). For an excellent discussion of the effects of practice on this test, see Whipple ('15 pp. 508-512). Special investigations by Starch, Lough, Kline and Munn are discussed. He reports also that Baldwin found that a group of Negro girls made only 62.4% as many substitutions and 245.3% as many errors as white girls did. Woolley and Fischer found a positive correlation between this test and school grades for children of both sexes, aged fourteen and fifteen.

Note: Substitution tests which involve only letters and digits are not included here.

References:

Woodworth and Wells '11
Lough '12
Pyle '13, '16, '19, '25
Woolley '15, '26
Whipple '14, (pp. 499-516)
Woolley and Fischer '14
Weidensall '16
Poffenberger '16
Pintner and Paterson '17
Pintner '18
Link '19
Dewey, Child and Ruml '20
Averill '24
Boody '26
Bronner, Healy, Lowe and Shimberg '27
Mohr and Gundlach '27
Davenport, unpublished mss.

No. 30 Visual Apprehensions Test

Devised by U. S. Public Health Service.

Date: 1917

Material: A board containing ten objects: gun, spool, fork, knife, cup, button, flower, doll, stone, toy horse.

Problem: After a short practice period on another board, to recall objects seen on the test board, which was exposed for four seconds.

Individual test. This test could be adapted to group testing

Directions: Verbal; pantomime by adaptation.

Responses: Verbal; adaptation to non-verbal form possible.

Standardization: This test was reported by Mullan, but no standardization was made. He tried it out on 291 literate and illiterate aliens, but found it unsatisfactory from the standpoint of differentiating the intelligent from the unintelligent.

Discussion: "Many intelligent immigrants, as shown by their excellent correlations in other tests, enumerated only two or three objects, while many less intelligent ones enumerated five or six objects." (Mullan '17, p. 33.)

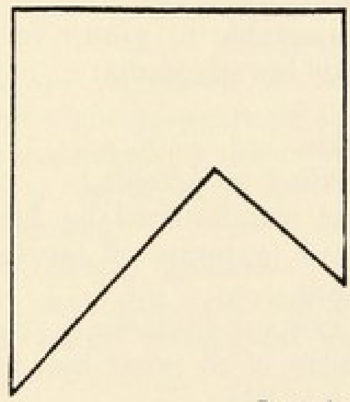
Reference:

Mullan '17

No. 31 Visual Designs

Devised by Binet; modified by the Army psychologists.
Date: 1900
Handled by Stoelting.
Price: $1.00
Material: A set of five cards, each containing a different design.
Problem: To reproduce the designs from memory.
Individual test: Design x is a demonstration design. Designs a and b single. c and d in pairs. This test could be adapted to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Time limit: One minute for each pair; ten seconds exposure.
Scoring: Points according to a scheme of credits. (Yoakum and Yerkes '20, p. 110 ff.)
Standardization: This test was part of the series given to the army.
Validity: r = .735 with Stanford-Binet mental age for 260 Army men.
r = .858 with Army Performance Scale for 260 Army men. (Yerkes '21, p. 404)
References:
Yoakum and Yerkes '20 Yerkes (Ed) '21

No. 32 Visual Designs Test



Reproduced by courtesy of Little, Brown & Co.

Devised by F. W. Ellis; modified by M. E. Goudge and Harry W. Crane.
Date: 1927
Published by Stoelting.
Price: $1.00
Material: A set of ten geometrical designs, graded as to difficulty. These are each on separate cards, and include the Binet-Simon designs.
Problem: After a brief exposure, to reproduce each design on paper.
Directions: Specific verbal.
Individual test. The test could be adapted to group testing.
Responses: Non-verbal.
Time: Five seconds for each exposure.
Scoring: Not established.
Discussion: Bronner, Healy, *et al*, consider these designs suitable for older

subjects. ('27, p. 205) Goudge reports that she has data on this test in preparation. (letter to the writers.) (See discussion on Test No. 33.)

References:

Terman '16 Goudge, (letter to writers).

Bronner, Healy, Lowe and Shimberg '27

No. 33 Visual Designs Test

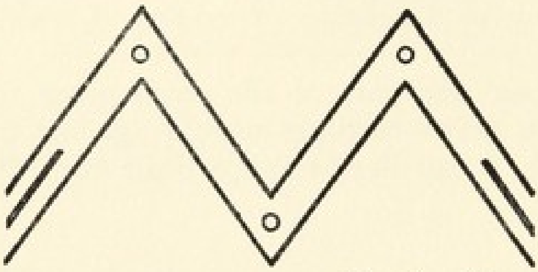
Devised by W. Healy.

Date: 1927

Published by Stoelting.

Price: Ten cents.

Material: A 2 x 4 inch card, on which is drawn a double-lined figure.



Reproduced by courtesy of Little, Brown & Co.

Problem: To reproduce from memory a line design which is exposed for five seconds.

Individual test. This test is adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Non-verbal.

Time: Exposure of five seconds.

Scoring: Sum of credits according to a table.

Standardization: This test was standardized by Meehan and Shimberg on 680 cases. Norms are expressed in terms of raw scores, for age and quartiles. Range: nine years to adult.

Reliability: r = .19 on 262 normal fourteen-year-olds. (Against Design I of Terman) As Bronner-Healy *et al* point out, these designs are all too easy for age fourteen. ('27, p. 204)

Discussion: Bronner, Healy, *et al* feel that success at the lower age levels or by duller subjects, is usually due to good visual memory. They point out that Terman thinks the most important feature of this kind of test is the measurement of the ability for rapid analysis. He thinks that there are too many lines for the ordinary memory span, and that these lines must be quickly grouped into some patterns for reproduction. (Bronner, Healy, Lowe and Shimberg, '27, p. 205)

References:

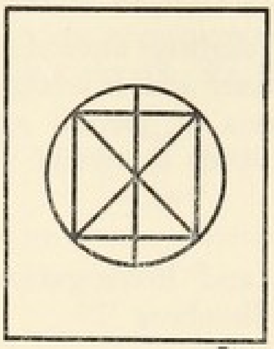
Terman '16 Bronner, Healy, Lowe and Shimberg, '27

No. 34 Visual Recognition of Forms Test

Devised by E. S. Jones.

Date: 1927

Published by Stoelting.
Price: $11.25 per hundred
Material: Two packs of cards, one containing fifteen, and the other twenty geometrical designs, ten of which can be identified in the first set; sample card.



Reproduced by courtesy of Little, Brown & Co.

Problem: After exposing first set, one at a time for five seconds each, Subject is required to sort second set on the basis of those previously seen or not seen.
Individual or group test.
Directions: Verbal; pantomime possibly.
Responses: Manipulative.
Time: Thirty minutes. (Jones)
Scoring: Positive and negative errors. Deduct ten for each from 100. Minimum score: —100.
Standardization: 1. Norms are expressed by Woolley on 358 males and 277 females, of seventeen and eighteen years of age, for scores and errors in percentiles. 2. Norms are expressed by Jones on 150 college students for time and in quartiles. (letter to the writers)
Reliability: r = .80 approximately.
Validity: r = .25 with scholarship. (Jones)
Discussion: Jones mentions that this test has caught some types of superior ability and feels that it would be a good test for superior adults. (letter to writers)
References:
Woolley '26 (pp. 147-151)
Bronner, Healy, Lowe and Shimberg '27
E. S. Jones, letter to writers.

No. 35 Visual Recognition of Pictures Test

Devised by G. M. Fernald.
Date: 1912
Published by Stoelting.
Price: $3.00
Material: Two groups of postcards, ten to a set; no duplicates, but pictures are closely matched.
Problem: After brief exposure, to recognize the first set of pictures when shuffled with others not previously seen.

Individual test. Test is adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal or non-verbal.

Scoring: 1. Percentage of correct recognitions noted. (2) Sum of errors in each set. (N. Y. Bureau)

Standardization: This test, in slightly altered form, was standardized by the New York Bureau of Analysis and Investigation on 885 cases. Norms are expressed for errors in terms of mental ages, based on 1911 revision of Binet-Simon tests. Range: seven to twelve years.

References:
Fernald '12
N. Y. Bureau of Analysis and Investigation '15
Bronner, Healy, Lowe and Shimberg '27

No. 36 Free Analogies Test (Pictorial)

Devised by S. C. Dodd; modified by P. C. Squires.

Date: 1926

Produced in the Psychological Laboratory, Princeton University.

Materials: Five sets of pictures are used in this test. A set comprises four blocks of pictures, four pictures to a block. Analogies board, upon which are mounted two boards, or platforms, whose edges are parallel with the length of the board. Four glass markers. (For details, see Squires '26, p. 116 ff.)

Problem: Given four blocks of a picture, to work out analogies between the pictures, under the condition that only one picture from each block may enter into any one analogy.

Individual test: Three practice analogies; five sets of test analogies. This test is not adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: Limit of two minutes for each set of analogies.

Scoring: Moves, successes.

Standardization: Squires experimented with this test on fifty Princeton seniors and fifty feeble-minded males from Vineland. Norms are not yet established.

Validity: $r = .11$ (successes) for the Princeton group, the criterion being the higher score on either the Thorndike or Princeton intelligence examination.

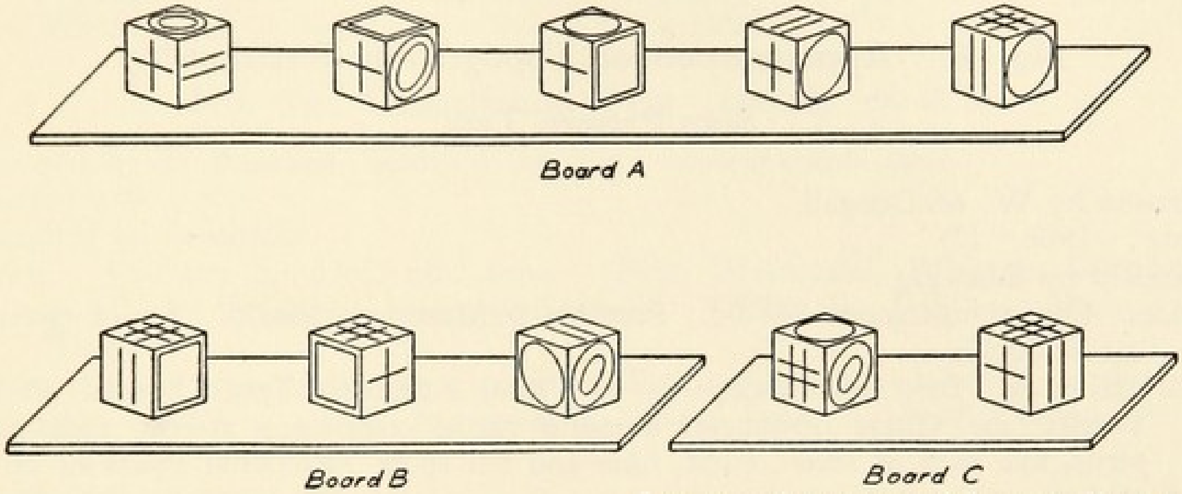
$r = .07$ for the Vineland group, criterion being Stanford-Binet mental age. (p. 184)

Discussion: "The free analogies test yielded the worst results of any one of the tests. This test should be administered to several age levels and to a large number of subjects in order to discover its real status." (Squires '26, p. 194)

References:
Squires '26

No. 37 Reasoning Test

Devised by P. C. Squires.
Date: 1926
Produced in the Psychological Laboratory, Princeton University.
Material: Platform board consisting of four platforms each 2 x 2 x ½″; two-inch cube; counter box with six compartments; six groups of counters; three "reference cube" boards. (For details, see Squires '26, p. 106 ff.)



Reproduced by courtesy of Princeton Univ. Press

Problem: A cube is shown with different geometrical designs on its surface, only three of which can be seen at one time. A single cube or column of cubes is then presented. The task is to determine what design is on the bottom face of the cube.
Individual test: Three practice items, five test items. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Time for entire item; moves; number of successes for sub-items of each test.
Standardization: Squires experimented with this test on fifty Princeton seniors and fifty feeble-minded males in Vineland. Norms are not established.
Validity: r = .07 (successes) and —.33 (time) for Princeton group, criterion being the higher score on either the Thorndike or the Princeton intelligence tests.

r = .18 (successes); and .19 (time) for the Vineland group, criterion being the Stanford-Binet mental age. (p. 184)

Discussion: "This test seems to be very difficult for all subjects so far tested with it." (Squires '26, p. 115).

"The test demands, among other processes, rapid re-adjustments on the part of the subject to sudden and relatively novel shifts in the test situation. Imagery processes are brought out rather well." (Squires '26, p. 168)

"With further investigation this test will be found to be a satisfactory one at the higher levels." (Squires '26, p. 193)

Reference: Squires '26

III

SENSORY TESTS

No. 38 Apprehension of Topographic Relationship

Spot-Pattern Test

Devised by W. McDougall.
Date: 1908 (?)
Handled by Stoelting
Price: Disc tachistoscope $85.00. Portable tachistoscope $64.75. Set of twenty cards, $0.40.
Material: A. Disc tachistoscope as in Visual Attention Test. Special set of twenty-nine square centimeter stimulus cards, comprising twenty spot patterns, five each of seven, eight, nine and ten spots. Prepared sheets of cross section paper; stop watch.

B. Special portable tachistoscope, consisting of a three-inch, focal plane photographic shutter, with wire release, mounted on a small easel adapted for holding stimulus cards; cross section paper as in A; seconds pendulum. (Whipple '14, p. 293).

Note: Langfeld and Allport have devised a cardboard tachistoscope which sells for $1.25 Dearborn and Langfeld have reconstructed this model in aluminum, which sells for $25.00. Stimulus cards $.60. Freeman also has devised two sets of cards at $3.30 and $.50, some containing figures, others words and others dots. (Stoelting)
Problem: To reproduce, upon cross-section paper, the arrangement of spots on the stimulus card.
Individual experiment; not adapted to group work.
Directions: Verbal, pantomime by adaptation.
Responses: Non-verbal.
Time: Each card is exposed two seconds; or 1.5 seconds, according to Goudge.
Scoring: Reported in the form of mean number of exposures needed to reproduce spot-paterns. (Goudge)
Standardization: This test, to the writers' knowledge, has not been used on large enough numbers to establish norms, but Whipple says that the test has a high correlation with estimated intelligence. He thinks that it may have diagnostic value, because of Burt's findings, in the ability of the test to differentiate groups. (Whipple '14, p. 295)
Reliability: $r = .50$ (Burt) $r = .90$ (Goudge for correlation between forms A and B of this test.
Validity: $r = .76 \pm .05$ P. E. and $.75 \pm .09$ P. E. against estimated intelligence (Burt)

Discussion: This test, as compared with the Visual Apprehension tests is supposed to test more directly the capacity for "locating the components of a complex visual stimulus" which was called "grouping" in the discussion under the latter tests. (Whipple '14, p. 290) Goudge's introspective records show that the patterns are always learned by a mental grouping of the spots.

References:

Whipple '14	Langfeld and Allport '16
Goudge '15	Squires '26
Freeman '16	Wallin '27

No. 39 Cancellation Test (usually of A's)

Devised by B. Bourdon; modified by many experimenters since.
Date: 1895
Handled by Stoelting.
Price: $.80 per hundred, pied type. (Pyle, Weidensall, Whipple, Woolley). $1.50 per hundred, geometrical forms. (Whipple, Whitley, Foster and Tinker).
Material: A page containing rows of small or capital printed letters, or symbols.
Problem: With a pencil, to cancel a certain letter or symbol, whenever it appears on the page.
Individual or group test.
Directions: Verbal, pantomime by adaptation.
Responses: Non-verbal.
Time: Varies with different experimenters.
Scoring: Usually the number of letters cancelled per minute plus a penalty for incorrect markings.
Standardization: This test has been tried out by so many different workers in so many different ways, that it is difficult to sum up the variations in any one statement. Reference should be made to the individual researches, some of which are reported below. Norms are offered by Pyle based on experiments with 592 boys and 628 girls. These norms are expressed in terms of age and sex, and show a sex difference in favor of the girls.
Reliability: Abelson (on dot-form test) r = .94 to .97. Simpson (regular "A" test) r = .60 to .72, and for his Geometrical Forms Test r = .69 to .91 (Whipple 14, p. 323)
Validity: Garrett and Lemmon find that there is usually a low correlation with intelligence and often lower than .40 with such tests as analogies, word building, completion, and the like. ('24, p. 424) For a discussion and summary of the earlier work on this test, see Whipple ('14, p. 305 f). He reports on coefficients of correlation with intelligence, as follows: an inverse correlation, r = — .40 between speed and class standing; and zero correlation for accuracy and class standing when one letter is cancelled, and r = .39 when four letters are cancelled. (Whipple) Wissler found r = .09 with class standing. (Wissler '01) Wyatt (er-test), found r's from .37 to .40; (anos test) .32 to .45; Burt, (oe-test) .39; Brown (er-test), 0 to .28; (anos test) .10 to .13; Simpson: "A" test with seventeen highly selected adults, .21; Abelson, dot form test with eighty-eight girls, .32 and, with forty-three

boys, .28; Descoeudres, with fourteen backward children, .67. See also Whipple's paragraphs on the correlation of the cancellation test with other tests. (Whipple '14, p. 324)

Discussion: Bourdon ('95) appears to have been the first user of the cancellation test. To him it was a test of discrimination. Binet ('99) used it to measure attention and adaptation. According to Whipple, "Oehrn ('96) proposed it under the title 'search for assigned letters' as a convenient test of attention for experimentation in individual psychology. Cattell and Farrand ('96) introduced it into the series of tests for Columbia University students in the form of the 'A-test' for 'rate of perception' and Thorndike ('05) employed it later, together with what may be regarded as variations of it (the 'a-t test', the 'e-r test' and the 'misspelled word test') for various comparative studies, particularly for his examination of the mental traits of twins." (Whipple '14, p. 305) Whipple cites the opinions of thirteen other psychologists as to the value of this test. He reports four variations: A, Cancellation of a single letter; B, Cancellation of more than one symbol; C, Cancellation of words, the a-t and the e-r test; D, Cancellation of misspelled words. He discusses the results of several investigations using one of these forms, particularly those of Woodworth and Wells ('11), Wissler ('01), Doll ('13), Whitely ('11) and Burt ('11). Whitley ('11) concluded that this test was significant of an individual's ability in visual perception. Simpson ('12) introduced a new note by providing a different type of material to be cancelled. With Garrett and Lemmon, "The recognition of words seems to be more closely related to the cancellation tests than the rapid motor test.". . . ."In so far as our tests do measure speed of marking and ability to find certain designated symbols, however, it seems reasonable to conclude that these two factors play an equal role in determining scores on the "A" test." ('24, p. 427) Bronner, Healy, *et al* feel that not only as a test of visual perception, but also in measuring an individual's power of sustained attention and the rapidity of his movements, the test belongs in the clinic. ('27, p. 224)

References:

Wissler '01
Whitley '11
Woodworth and Wells '11
Simpson '12
Doll '13
Pyle '13 and '16
Hollingworth '14
Whipple '14 (pp. 305-329)
Woolley and Fisher '14
Weidensall '16
N. Y. Bureau of Analysis and Investigation '15 and '17
Pintner '18
F. L. Wells '19
Morganthau '22
Garrett and Lemmon '24
Woolley '26
Bronner, Healy, Lowe and Shimberg '27
Smith '27

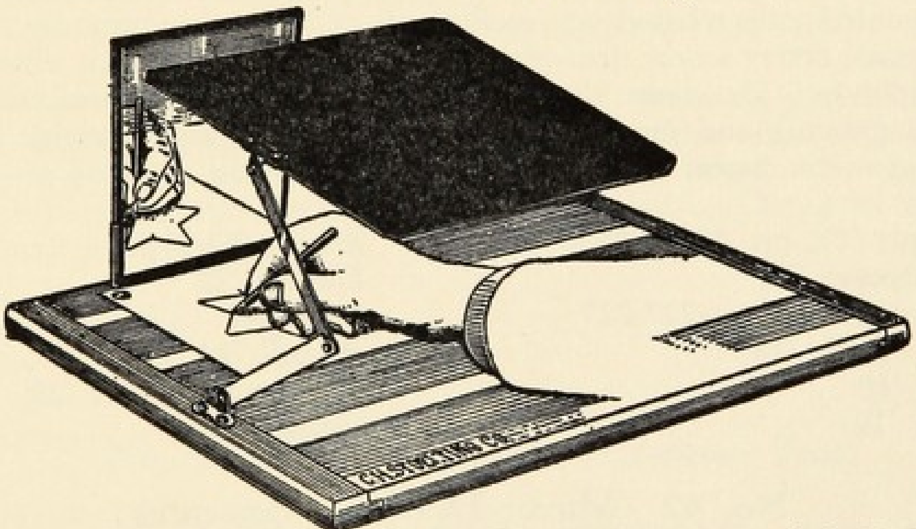
No. 40 Dot Counting Test

Devised by A. Binet; adapted by H. A. Knox.
Date: 1899
Handled by Stoelting.

Price: \$1.85. \$0.15 for Knox's test (card containing sixty dots).
Material: Two duplicate sets of thirty-one printed test cards; stop watch.
Problem: To count as quickly as possible the number of dots which are arranged in an irregular group or in lines of varying length and spacing.
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Verbal; manipulative by adaptation.
Time: No time limit; Subject's time should be recorded.
Scoring: Errors are counted by subtracting the number given from the true number.
Standardization: This test has evidently not been standardized to measure intelligence. Smith has included it in his scale. ('27, p. 190)
Discussion: Binet concluded that success in this test depended more on care than on intelligence. However, intelligent children are found to make fewer errors than do dull children. Winteler found a difference between the performance of bright and dull children as groups, in which both speed and accuracy favored the bright. Whipple comments that these findings suggest that the "tendency to inverse relation of speed and accuracy may not be so clearly evident as in many other tests. (Whipple '14, p. 332)
References:

Binet '99
Whipple '14 (pp. 329-332)
Knox '14
Mullan '17
Smith '27

No. 41 Mirror Drawing



Reproduced by courtesy of C. H. Stoelting Co.

Devised by G. M. Whipple
Handled by Stoelting
Price: Apparatus \$6.25. Cards: Six-pointed cardboard star, \$2.25; paper \$1.00. Six cardboard forms with figures graded in difficulty, printed in red, \$6.00; paper, \$3.00. Strong, Pyle and Foster have also devised paper tracing forms, priced \$1.00, and \$2.00 respectively. Scoring stencil (Israel's) of celluloid, \$1.00.

Material: Two kinds of diagrams printed in red ink, for tracing (a) six-pointed star, (b) a set of six patterns, each based upon a group of twelve points arranged at equidistant intervals in a circle about its central point, with guiding lines joining the thirteen points in irregular fashion; mirror: cardboard screen. (Whipple '15, p. 486)

Problem: To follow the design with a pencil while looking at the mirror reflection.

Time: No time limit, but time is taken.

Scoring: Six trials allowed. Subjects may be compared with respect to (1) their first trial, (2) their sixth trial, (3) average of all six trials or their rate of improvement in per cent of gain, either (4) in the sixth, compared with the first trial, or (5) in the average of the last three, compared with the average of the first three trials. (Whipple '15, p. 489)

Standardization: Whipple reports that Yoakum and Calfee have supplied norms for the patterns. Burt obtained a correlation between speed and estimated intelligence of .67 P. E. ± .07 for elementary school boys and of .54 P. E. ± .14 for preparatory school boys. Burt and Moore found r = .60. However, Calfee found approximately zero correlation in her group of elementary school children. (Whipple '15, p. 494)

Reliability: r = .52 (Burt and Moore, according to Whipple)

Discussion: Whipple reports that girls are better than boys on this test. The effects of practice are paramount. (See Whipple's discussion '15, p. 491-494) For a historical sketch of psychologists' interest in mirror writing, (See Whipple '15, p. 486) Cooke '18 found this test of use in showing traits of nervousness, carefulness, persistency or lack of concentration. He doubted any relation to intelligence. Clinton has recently found that boys make better scores than girls up to the age of thirteen; after that girls outstrip boys. He states also that there is no positive correlation between mirror drawing and intelligence although there is a progressive age to age increase in the motor ability required. (Clinton '30, p. 228)

References:

Whipple '15, pp. 485-498
Poffenberger '16
Weidensall '16 pp. 222-227
G. R. Wells '18
Cooke '18
Perrin '19
Strong '22
Hunter '23
Pyle '24
Averill '24
Foster and Tinker '29
Clinton '30.

No. 42 Mirror Drawing Apparatus

Devised by G. M. Whipple; modified by G. R. Wells
Manufactured by Stoelting.
Price: $60.00

Material: Two concentric stars of metal, the outside line of the smaller being smaller than the inside line of the larger. These are connected electrically with each other, to an electric counter and to a metal stylus; also a screen and a mirror.

Problem: Subject must keep the stylus within the inscribed figure. Each time the stylus moves from the inscribed star the counter registers.
Individual experiment; not adapted to take care of more than one at a time.
Directions: Verbal; pantomime possibly.
Responses: Manipulative.
Time: Thirty seconds.
Scoring: Automatic scoring on mechanical counter.
Standardization: No norms or standardization reported.
Discussion: This test, according to G. R. Wells, probably tests "motor control, rate of learning of this particular process, etc." (letter to writers)
References:
Whipple '15, pp. 485-498
G. R. Wells '17 and '18, and letter to writers.

N.B.: A mirror drawing apparatus called a Stabilimeter was devised by G. S. Snoddy to study the learning curve under various conditions of practice. This can be obtained from Indiana University for $20.00 (less the electric counter. Standards for once-a-day series and one minute interval practice conditions will be supplied without charge). These are for adults. (Ref. G. R. Wells, letter to writers). The Stoelting prices for the Snoddy apparatus are: Apparatus I, without recording apparatus, $80.00; Apparatus II, without recording apparatus, $85.00; Apparatus III, without recording apparatus, $45.00.

Gopaleswami's modification of Snoddy's apparatus is priced by Stoelting at $82.00. Instead of having a piece of glass for a base, as in Snoddy's, this has a piece of tracing cloth under the star pattern which is held stretched over electrical contacts. A sheet of paper and a sheet of carbon are placed between the plate and the star pattern, so that a record can be made of the stylus movements.

Gopaleswami's Mirror Groove Apparatus has the same general arrangement as the drawing test except that it presents only the upper half of a star pattern, with a narrow groove cut into it, along the side of which run blind alleys at different turnings of the design. These blind alleys are listed as 10 habits, 8 opposites, and 8 variables. Price: $76.00 Stoelting)
Reference:
Gopaleswami '24
Snoddy '20, 26

No. 43 Range of Visual Attention Test

Devised by G. M. Whipple, after earlier tachistoscopes. (See Whipple '14, p. 264); modified by several experimenters.
Date: 1907 (?)
Handled by Stoelting
Price: Tachistoscope $85.00
Material: Disc tachistoscope; frosted tabular lamp, 16 C. P.; two four-inch clamps; blanks of cardboard nine centimeters square; two complete sets of Wilson's gummed black letters and figures, size three; drawing ink and

ruling pen; head-rest; fifty vibrations dry-contact fork; dry battery; connecting wire.

Problem: To report "orally or preferably by drawing" what is seen through tachistoscope exposure. The cards have been prepared by pasting on them letters, digits, spots, lines, geometrical figures, and the like, in varied fashion.

Individual experiment; not adaptable to group work.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal and non-verbal.

Time: Point exposure of sixty sigma for each card.

Scoring: In general, the number of concrete objects correctly reproduced.

Standardization: Whipple found that "when a series of *unrelated objects* is exposed, the average number of impressions that can be grasped in a single exposure lies between four and five." For a discussion of early experimental work in this field, see his excellent summary. (Whipple '14, p. 269 ff) The relation between range of visual attention and mental ability is uncertain, (Whipple '14, p. 275. See also Squires '26)

Discussion: Whipple discusses at length the special problems involved in tachistoscope experiments, dealing largely with the psychology of reading. He writes: "Practice has a curiously small effect upon the range of attention, when once the period of preliminary habituation to the arrangement of apparatus and method is passed." (Cattell '86, Hylan '03) Also: "There is only questionable evidence between the range of visual attention and mental ability. (Whipple '14, p. 273 f)

References:

Whipple '10, '14
Freeman '16

No. 44 Tapping Test

Devised by G. M. Whipple.

Handled by Stoelting.

Price: $10.00

Material: Tapping board, 55 x 10 centimeters, with brass 10 centimeters square on either end; tapping stylus with connecting wire attached; the kymograph with accessories; double-time marker; seconds pendulum; support with leveling screw and right-angle piece to hold time marker; table clamps for tapping board; large sheet of gray or white cardboard; two short-circuiting keys; stop watch; four dry cells, flexible covered wire. (Whipple '14, p. 131)

N.B.: The Whipple-Healy tapping test material consists of a piece of paper divided into 150 half-inch squares. This is offered as a test for motor-coordination.

Problem: To tap the plate with the stylus as rapidly as possible within the time limits given.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Time limit ranges from five seconds (Binet and Vaschide) to two minutes (Thompson), or to completion of 100 taps (Dressler). Whipple gives thirty seconds as being adequate for ordinary purposes.

Scoring: The "total efficiency" of a "record" (five trials with the same hand) is the average of the sum of the taps per trial. (For indexes of fatigue and handedness, see Whipple '14, p. 141 f).

Standardization: For a sample record, see Wells '08. For practised adults, the initial rate is about 8.5 taps per second, with a range from five to fourteen taps. (Whipple '14, p. 138) The correlation between tapping ability and mental and social status is found to be generally positive. Smedley, Gilbert, Bolton, Kirkpatrick and Burt obtained coefficients ranging from .41 to .65 and Abelson got r = .28 to .42. Bagley and Whipple found the correlation to be indifferent; and Binet and Vaschide found a positive correlation with twelve-year-olds and an inverse one for sixteen to twenty-year-olds. Bolton also found that "good" children showed a greater practice improvement than did poor ones. Whipple points out that the constant individual differences found by the test are undoubtedly conditioned in a general way by fundamental neural factors. The rate of tapping increases with age between six and eighteen years, and boys are generally faster than girls. The effect of practice is a gradual speeding up of tapping ability. For data on 'warming up, 'spurts', and so forth, see Whipple. ('14, p. 140)

Reliability: r = .51 (Burt); r = .92 on girls, and r = .91 on boys (Abelson) (Whipple '14, p. 144)

Discussion: The tapping test has been used in many forms, such as marking or making dots with a pencil (Binet and Vaschide), puncturing a small square of paper with a pointed stylus (Healy), and trilling upon telegraph keys (Abelson, Burt).

References:

Healy and Fernald '11
Whipple '14 (pp. 130-147)
N. Y. Bureau of Analysis and Investigation '15
Dunlap, K. '21
Hewes '22
Hollingworth and Monahan '26
Bronner, Healy, Lowe and Shimberg '27
Mohr and Gundlach '27

No. 45 Visual Apprehension Test A

Devised by G. M. Whipple, after Aiken ('96)
Date: 1907 (?)
Handled by Stoelting
Price: Tachistoscope $85.00

Material: Same as in the Range of Visual Attention Test, (No. 43) except that cardboard blanks are replaced by blanks of stiff paper, 12.5 x 20 centimeters, and that the exposure card holder and the opening in the screen of the instrument are adjusted to these dimensions. Metronome with electrical contacts or double time marker on seconds' pendulum. Whipple suggests the following type of exposure cards: (a) groups of irregularly arranged dots or circles; (b) cards on which are pasted pictures cut from old magazines; (c) single line of eight to ten three-letter nonsense syllables; (d) meaningless drawings; (e) typewritten four-line stanzas of simple but not well known poem; (f) columns of digits. (Whipple '14, p. 279 f)

Problem: To report what is seen through a period of several seconds of tachistoscopic exposure. For dots, a statement of their number, supplemented possibly by a sketch of their position; for nonsense syllables, poetry or digits, an oral or written report; for drawings, a pencil sketch; for pictures, a verbal description supplemented by a pencil sketch.

Individual experiment; group adaptation possible.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal and non-verbal.

Time: Exposure for each card of three seconds.

Scoring: (a) Number of dots correctly counted; (b) pictorial reproduction graded on a scale of ten; (c) each letter in nonsense syllables, correctly reported sets one unit; (d) drawings rated on scale of ten; (e) one unit for each word of poetry correctly reproduced; (f) one unit for each digit correctly reported; .5 deducted for each error of transportation or insertion in the case of (c), (e) and (f). (Whipple '14, p. 282)

For Discussion and References, see Visual Apprehensions Test B. (No. 46) Whipple '14, (pp. 278-283)

No. 46 Visual Apprehension Test B

Devised by G. M. Whipple

Date: 1907 (?)

No manufactured material needed, except pendulum which is handled by Stoelting.

Price: $17.50

Material: Small table; table cover, preferably gray; seconds pendulum; piece of cardboard about 30 x 45 centimeters, and a sheet, 22 x 28 inches, of gray cardboard; collection of miscellaneous small familiar objects, i. e., pencil, spoon, leaf, key, etc. Ten different objects will be needed at each exposure. (Whipple '14, p. 283)

Note: Franz and Tinker have devised stimulus cards for use with, or without, a tachistoscope. Tinker's $16:50; Franz's $0.50, Stoelting.

Problem: After a short exposure period to enumerate as many objects as possible.

Individual experiment; not adapted to group work.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal or non-verbal.

Time: Six seconds for each exposure.

Scoring: One unit's credit for each object named.

Discussion: Whipple's experiments bring out the small increase in range of visual perception in comparison with range of attention. Thus, four or five objects are grasped during an exposure of ten to fifty sigma, and only six objects during an exposure one hundred times as long. Individual differences in capacity for quick apprehension are clear, some subjects excelling in estimation of dots, others in reproducing drawings. Whipple analyzes the data secured in his experiments as showing that efficiency in visual apprehension depends on: (1) native capacity for concentration; (2) degree of attention during the exposure; (3) capacity to assimilate particular types of material; (4) ease of assimilation of particular test card in use; (5) obstruction or distraction; (6) ideational type, visual or auditory-minded subjects; (7) restriction of attention to what is within subject's ability to grasp; (8) grouping of constituent elements in exposure field. (Whipple '14, p. 288 f.)

References:

Whipple '14, (pp. 283-290)
Franz '12, '19
Cornell '17
Mulhall '17
Foster and Tinker '23

IV

SORTING TESTS

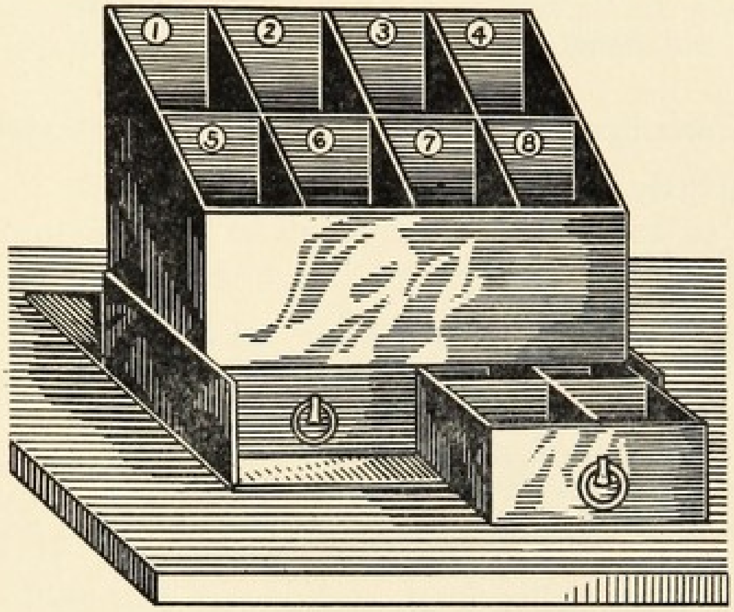
No. 47 Ball-Tossing Apparatus

Devised by W. H. Pyle.
Date: 1923
Handled by Stoelting.
Price: $11.00 for apparatus; $4.50 for fifty extra balls.
Material: A bag, about six inches in diameter, suspended from a wire ring attached to 9-inch wooden uprights; fifty rubber balls about 15-16 inch in diameter.
Problem: To toss the balls into the bag. Eight series of trials are allowed, four on first day and four on next.
Individual experiment only.
Directions: Verbal or pantomime.
Responses: Manipulative.
Scoring: Number right.
Standardization: Pyle offers the actual scores of twenty-four subjects over a two-day period, and the scores of one subject over a hundred day period. (Pyle '23, p. 35 ff) Otherwise norms are not established, nor has this material been standardized as a test of intelligence.
Discussion: Pyle thinks that the most important part of this experiment is its qualitative aspects. Delicate co-ordinations are required; greater steadiness is acquired with practice; 'runs" of two or three consecutive successes are encountered; and so on. (Pyle '23, p. 37 f)
References:
Pyle '23

No. 48 Card Sorting

Devised by J. Jastrow; modified by many experimenters.
Date: 1898
Handled by Stoelting.
Price: Jastrow's, $45.00 to $105.00; Woolley, $15.00; Link, $6.25; Minnesota series, $7.25; Pyle's, including 150 cards, $52.00; Averill's, $10.00; Kline's, $108.00.
Material: Card sorting test materials vary with the particular revision used. Jastrow's original apparatus consisted of two groups of four boxes. This was developed into an eight-compartment box, adjustable into two parts. Woolley and Fischer's apparatus was a modification of Jastrow's, evidently one of the two adjustable parts, and consisted of "a black box with four compartments, each marked with a circle of color; blue, green, yellow, red.

A pack of forty-eight cards, each marked with a circle of one of the colors, twelve of each kind; order arranged." (Woolley and Fischer '14) For a description of Jastrow's cards, see Jastrow, ('98, 279 f) Pyle ('25) used seventy-five cards and fifteen compartments. Link ('19) used sample sets of six and eight cards and test sets of fifty cards, upon which letters were placed. He also used a sorting board. A sorting test is included with the Minnesota series of mechanical ability tests, consisting of two wooden trays, and playing cards, half with blue, half with red backs. The Klines ('27) have devised a case for testing the memory of position in a vertical plane. It contains fifty-four pigeon-holes, which are slightly tilted. (See Test No. 51)



Reproduced by courtesy of C. H. Stoelting Co.

Problem: To sort cards as quickly as possible according to directions: color, form, etc. or a combination of both.

Individual or group test.

Directions: Verbal, or pantomime by adaptation.

Responses: Manipulative.

Time: Usually no time limit.

Scoring: Usually time.

Standardization: This test was reported on by Whitley; also by Woolley and Fischer for 750 subjects; and by Pyle for 1326 males and 1597 females. Pyle offers norms for age and sex. Range: from eight to eighteen years ('25, p. 16) Dewey, Child and Ruml ('20) report scores for time and accuracy for girls and boys, aged nine to thirteen.

Reliability: Whitley, who studied intensively certain tests of individual differences, reports correlations of .99 for shape, .52 for color, .61 for size, .63 for suit, .43 for six numbers, all with the average of these five tests. (Whitley '11, p. 83)

Validity: Burt reports a correlation of .52 to .56 with teachers' estimates of intelligence. ('09, p. 137)

Discussion: This test, like the cancellation test, more properly belongs to the psycho-physical group than to the mental test group. It also has been claimed to test various things, but Pyle offers norms for intelligence. As Woolley and Fischer ('14, p. 101) and Weidensall ('16, p. 82) point out, it is probably a test for "eye-hand coordination".

References:

Jastrow '98	Averill '24
Burt '09	Baldwin and Stecher '25
Whitley '11	Woolley '26
Woolley and Fischer '14	Kline and Kline '27
Weidensall '16	Bronner, Healy, Lowe and Shimberg '27
Pyle '16, '25, '28	L. D. Anderson '28
Link '19	Foster and Tinker '29.
Dewey, Child and Ruml '20	

No. 49 Complicated Motor Learning (Manthanometer)

Devised by W. H. Pyle.
Date: 1923
Manufactured by Stoelting.
Price: $90.00 (Boynton's $150.00)
Material: Ranschburg apparatus, telegraph key, color disc, six dry batteries; stop watch; manthanometer; 24 large and 24 small marbles in four colors; a four-section tray on top of manthanometer.
Problem: Subject must sort colored balls, and balls of two sizes simultaneously, according to directions presented through the slip of the Ranschburg apparatus. This involves simultaneous movements of both hands, and movement of both feet also involving a high degree of concentration. The subject presses a key to get stimulus (directions), and presses foot pedals for *different compartments,* and the like.
Individual experiment; too complicated for group work.
Directions: Verbal.
Responses: Manipulative.
Time: Thirty minutes.
Scoring: Time in seconds; number of errors. (Pyle '23) Efficiency score = 96 — total numbers of errors x 60 ÷ time in seconds. (Pyle '25, p. 58 f)
Standardization: Pyle, in 1923, offered scores made by ten students on each of thirty-one trials over a period of a week. (Pyle '23, p. 71 ff) Later he offered scores made by children for one or two sortings, which show a progressive age increase. These scores are distributed for age and sex, ranging in age from nine to seventeen years, and favoring the girls up to sixteen years. (Pyle '25, p. 56 f)
Discussion: This is an attempt to measure complicated sensori-motor learning and concentration. It throws light on the nature of the learning curve, the relation of speed to accuracy, and individual differences in keeping several processes going simultaneously. Pyle stresses the importance of noting the qualitative aspects of the reaction method in learning and remembering the different associations and movements; which things were easiest to learn, and which most difficult, type of attention given, types of reactions to instructions

and so on. ('23, p. 71 and 79) Pyle thinks that the experiment measures, in addition to general learning capacity, a specific type of ability, such as that required to operate an automobile.

References:

Pyle '23, '25, and '28

No. 50 Cube Distribution Test

Devised by J. C. Chapman.

Date: 1921

Material: Four small boxes painted red, blue, yellow and green; and ten cubes.

Problem: Cubes to be distributed according to directions.

Individual test; adaptable to group testing.

Directions: Verbal

Responses: Manipulative.

Standardization: This test has apparently been discarded; it was reported by the Army psychologists in the Memoirs of '21, but no norms were given other than a table showing percentages of passes for five different intelligence quotient groups.

References:

Yerkes (Ed.) '21, p. 364.

No. 51 Distributing Case

Devised by L. W. Kline

Date: 1927

Handled by Stoelting.

Price: $108.00

Material: a. Wooden case, 52 x 92 x 11 centimeters, divided into 54 pigeonholes, the floors of which are slightly tilted; curtain in front; clip to each pigeonhole for insertion of label. b. Charts A and B, showing the plan to be followed in sorting. c. Ten decks of playing cards. d. Blanks for entering record. (p. 220 f)

Problem:

1. After a preliminary practice period—until maximum speed is attained—to sort the cards into the case according to a specified plan.
2. After relabeling the pigeonholes, to resort the cards according to a new plan.

Individual experiment; not adapted to group work.

Directions: Verbal.

Responses: Manipulative.

Time: Indefinite, except for number of trials allowed in practice period.

Scoring: Time; errors; and note also jerkiness in overcoming inhibition.

Standardization: This material has not been standardized, but is included here because of its promise of usefulness in devising experiments and tests on learning ability, overcoming associative inhibitions, and other adaptations of the learning situation.

Discussion: This material offers the experimenter an opportunity to set up a number of devices, exercises, tests, and so forth, on the acquisition of motor skill, inhibition, transfer and associative learning. (p. 223)

References: L. W. Kline and F. L. Kline '27

No. 52 Marble Sorting

Devised by W. H. Pyle.
Date: 1925
Handled by Stoelting.
Price: $26.50
Material: Box with nine compartments, each of which has an uncolored picture of an animal on it; a box of ninety colored animals; a key showing pictures of nine animals in colors corresponding to the nine colors of the marbles.
Problem: To sort the marbles as quickly as possible, each to be matched with the animal having its corresponding color, and then to be put in the hole by its picture.
Individual or group test.
Directions: Verbal; pantomime possibly.
Responses: Manipulative.
Standardization: This test was standardized by Pyle on 639 males and 652 females. Average scores and average errors are given. These norms are expressed in terms of age and sex. Range; from six to twelve years. ('25, p. 37 ff)
Discussion: The marble sorting test can be given to children or to others who cannot read or who do not know numbers.
References: Pyle '25, '28 Bronner, Healy, Lowe and Shimberg '27

No. 53 Perception Test

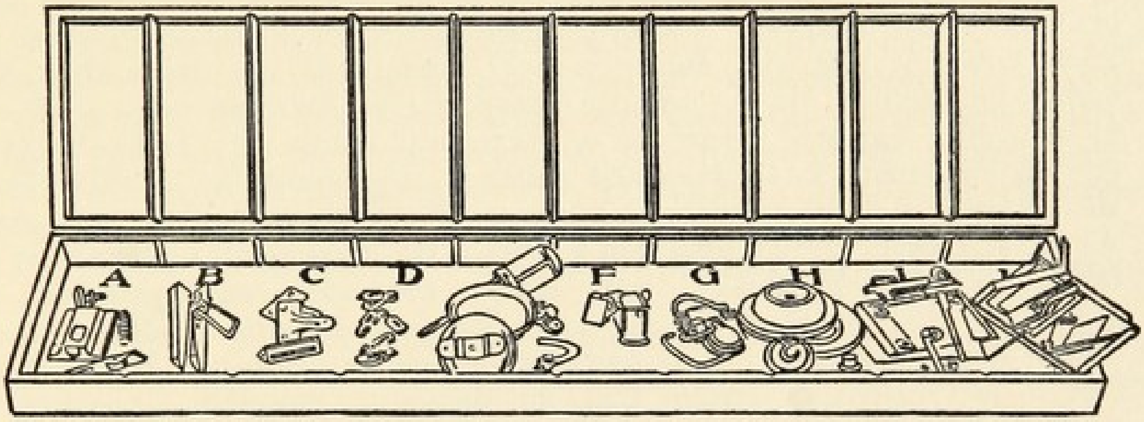
Devised by C. B. Cornell.
Date: 1917
Material: A set of fifty cards, playing card size, five different kinds, ten of each kind; three colors: blue, green, red; and three forms: circle, square and triangle. Each card has three spots, blue at the top, green at the center and red at the bottom, the variation in the different sets being produced by interchanging the forms.
Problem: To sort the cards according to assigned similarities.
Individual test; adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: No time limit.
Scoring: Time, number of piles, and errors.
Standardization: This test was standardized by Cornell on 550 public school children. Norms are expressed in terms of age averages. Range: six to fourteen years.
References: C. B. Cornell '17

V

SPECIAL ABILITY TESTS AND MECHANICAL PUZZLES

No. 54 Assembly Test of General Mechanical Ability

Devised by J. L. Stenquist.
Date: 1923
Handled by Stoelting
Price: \$12.50 for each series. Manual \$0.40. Scoring blanks \$0.80 per hundred.



Reproduced by courtesy of Little, Brown & Co.

Material: Three hinged boxes, 5½ x 2¼ x 24¼ inches, each containing ten unassembled mechanical devices. Series I and II are of equivalent difficulty and serve as alternates for Grade V and over. Series III, somewhat easier. is for Grades III, IV, V and VI.
Individual or group test.
Directions: Verbal; pantomime possible.
Responses: Manipulative.
Time: Thirty minutes.
Scoring: Plus or minus partial score values are allowed for various degrees of perfection in each model. These are converted into T-scores according to a table. A short form of scoring is also offered. (Stenquist '23, pp, 96–98)
Standardization: This test was standardized by Stenquist on 452 and 459 public school children from Grades IV to high school level. The test was also tried out rather extensively in the army. Stenquist offers grade norms for what they are worth, age norms and T-scores which permit of direct comparison with T-scale scores of other tests.
Reliability: "It is probable that the true reliability is between .6 and .7. For two groups, the high school class and the seventh and eighth grade boys, it runs up even higher. This degree of reliability is probably as high as can generally be obtained with such material, but it is not all that could be desired. It is to be hoped that further experimentation will result in scales of higher reliability." (Stenquist '23, p. 57)

Validity: r's range from .00 to .51 with tests of "general intelligence" as applied to Army men, public school children and feeble-minded children. Stenquist found $r = .23 \pm .04$ for Series I and $.338 \pm .06$ for Series II with intelligence. ('23, p. 73) r's are .24 for men and .13 for women on a three-hour intelligence test as applied to 82 adults. (Toops '23, p. 24)

r's range from .42 to .90 with ratings of two shop teachers. (Stenquist '23, p. 58 f) When correlated with another mechanical ability test, I. E. R. Assembly, on a group of fifteen "special class" girls, $r = .80$ (Toops '23, p. 46).

Discussion: This test was given to 14,610 Army men and found to correlate but little with "general intelligence". (Yerkes Ed. '21, p. 321) Stenquist states that his test "selects ability markedly different from that discovered by verbal tests of general intelligence." He found that the test detected those qualities which make for success in shop and manual training classes. He adds, however, that this ability is not "trade skill any more than it is verbal intelligence. It is rather a composite of common sense and skill in managing physical objects of a mechanical nature. It might be called general mechanical intelligence and ability." (Stenquist '23, p. 60) Toops also found a low correlation between this test and scores on Army Alpha and other group intelligence tests. He recognizes a distinction between "ability with things" and "ability with ideas and symbols." He found also that the Stenquist Assembly Test was a better predictor of success in shop work than was a composite of three paper tests. He concluded that it is "reasonably certain that the Stenquist Assembly is to-date the most important single test contribution to the measurement of general mechanical ability." (Toops '23, p. 23-27)

References:

Yoakum and Yerkes, '20
Yerkes (Ed.) '21
Stenquist, '23
Toops, '23
Dougherty '26
Peterson and Lanier '29

No. 55 I. E. R. Assembly Test for Girls

Devised by H. A. Toops
Date: 1923
Handled by Stoelting.
Price: \$6.70; scoring blanks, \$1.50 per hundred.
Material: A box containing material for eleven test items; A, stringing beads; B, inserting tape; C, rosette; D, cross-stitch; E, key-ring; F, clip-chain; G, tape-sewing; H, trunk-tag; I, card-wrapping; J, booklet; K, trimming paper.
Problem: Eleven projects as indicated in above summary of material.
Individual or group test.
Directions: Simple verbal; or pantomime by adaptation.
Responses: Manipulative.
Time: Forty-five minutes.
Scoring: Partial credits according to scoring sheets.
Standardization: This test has "to-date not been adequately standardized, nor is its reliability known." (Bronner, Healy, Lowe and Shimberg, '27 p. 208)

Reliability: No information.

Validity: Toops considers that the I. E. R. Assembly test "measures an ability less closely allied to ability with ideas and to success in school work than to the ability measured by the low level clerical tests and the Stenquist Assembly Test. It seems to do for the girls what the Stenquist Assembly does for the boys, but not so clearly and emphatically." (Toops, '23, p. 46) When given to 318 public school girls the average of four correlations was: r = .32 with Arith.—Re. intelligence (p. 22, insert)

Toops remarks, that the "I. E. R. Assembly Test does not depend upon intelligence in the case of boys; and that it does depend upon the abilities measured by other mechanical tests, about as highly as these other tests depend upon each other." ('23, p. 44)

Later: "The I. E. R. Assembly test is known to depend more upon intelligence than does the Stenquist Assembly Test as determined by the correlations respectively with intelligence." r = .80 with Stenquist Assembly on a group of 15 girls in a 'special class.' ages above 11. When administered to 318 girls, twelve to fifteen years old, the following correlations were obtained with Stenquist Assembly: r = .48, .47, .42, .31 (p. 22.) His evidence "points to a marked increase in relationship between the two tests due to the practice in mechanical things to which this class has been subjected." (Toops '23, p. 46)

Discussion: "The I. E. R. Assembly represents a clever attempt to supply a test for girls analogous to the Mechanical Assembly for boys." (Bronner, Healy, Lowe and Shimberg '27, p. 208)

References:

Toops, '23 Bronner, Healy, Lowe, Shimberg '27

No. 56 Instruction Box

(B. H. L. S. No. 112)

Devised by J. W. Hayes; modified by W. Healy.

Date: 1911 (?)

Handled by Stoelting.

Price: $35.00

Material: Wooden box about 6 x 5½ x 3½ inches.

Problem: To follow instructions which will result in opening the box.

Individual test; three trials if necessary. This test is not adaptable to group testing.

Directions: Verbal; pantomime doubtful.

Responses: Manipulative.

Scoring: Total time and steps in performance.

Standardization: No norms have been established.

Discussion: Bronner, Healy *et al*, representing the makers of this box, state that it was devised to show the Subject's ability to follow directions while working with concrete material. They point out its parallelism with every-day situations, such as opening a safe, a cash register and the like. The situation is more true to life than is to be obtained in the more formal verbal directions tests, because, in this case, the prospect of opening the box acts as an incentive to action. (Bronner, Healy, Lowe and Shimberg, '27, p. 232)

References:

Healy and Fernald '11
Healy '15
Bronner, Healy, Lowe and Shimberg '27

No. 57 Instruction Box
(B. H. L. S. No. 42)

Devised by J. W. Hayes.
Date: 1920
Handled by Stoelting.
Price: $48.75
Material: Instruction box; diagram of instructions necessary to open the box.
Problem: By manipulating the knobs and levers to open the box; six operations to be carried out.
Individual test; three trials may be given. The test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring: Number of trials necessary to open the box.
Standardization: This test was used by Woolley ('26) on a group of seventeen-year-olds, 464 males and 361 females. Her scores are distributed by sex and in percentiles.
Reliability: Woolley found this test "unsatisfactory from the point of view of distribution of the measure." ('26 p. 146)
Discussion: There is still another instruction box reported by Dewey, Child and Ruml, who used it and found it unsatisfactory. They advised a more complex mechanical construction in which each move would depend on the preceding one so that performance would be blocked by errors. ('20 p. 31–35)
References:

Dewey, Child and Ruml, '20
Woolley, '26
Bronner, Healy, Lowe and Shimberg, '27
Hayes, letter to writers

No. 58 Knot-Making Test

Devised by H. A. Ruger; modified by M. M. Mandl.
Date: Reported by Steacy 1919.
Material: Fifteen folders, upon the inside covers of which are drawings of the knots to be made; fifteen loose pieces of cardboard, 10½ x 7¼ inches. To these cards are attached loose pieces of sash cord, one piece if the knot is to be made single, and two pieces if the knot is to be made double. The knots to be made are arranged in their order of difficulty.
Problem: To make with the piece of card, a knot like the one in the diagram.
Individual test: Two forms; fifteen different knots. This test is adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Thirty minutes.
Scoring: Points according to a table of values. (Steacy '19, p. 8)

Standardization: Steacy tried this test out experimentally on thirty-one boys and thirty-eight girls in the sixth grade of a New York public school, and distributed his scores for sex Otherwise the test is not standardized.
Reliability: r = .65 for one series with other.
Validity: r = .50 with other tests in Steacy's battery, excluding handwriting. (Steacy '19, p. 38)
Discussion: Steacy says that this test "requires synthetic ability. The subject must be able to understand a pattern and to apply the pattern to concrete substances. He needs also some power of visual imagery and some manual dexterity." (Steacy '19, p. 38)
Reference:
Steacy '19, p. 7
Ruger, personal interview

No. 59 MacQuarrie Test of Mechanical Ability

Devised by T. W. MacQuarrie.
Date: 1925
Published by the Research Service Co.
Price: $1.50 per twenty-five, complete with directions, scoring keys, available norms, etc.
Material: Seven parts, printed in booklet form.
Problem: Seven tasks: 1, tracing; 2, dotting; 3, location; 4, tapping; 5, copying; 6, blocks; 7, pursuit.
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal; pencil and paper.
Time: Varies from thirty seconds to two and one-half minutes for the different tasks.
Scoring: Usually the number correct in each task.
Standardization: This test was studied by MacQuarrie who, however, does not report the ages or grade location of his Subjects. ('27, p. 336 f.) A more careful report has been prepared by Kefauver who was interested in comparing the diagnostic value of the MacQuarrie, the Stenquist Mechanical Aptitude Tests and the Terman Group Test in predicting the success of pupils in machine, electrical, automobile, mill-cabinet, and forge shops. His Subjects were 101 high school students, divided into small groups for shop instruction. Scaled rankings in shop success were supplied by their teachers.
Reliability: r = .90 for the whole test, on groups of 35, 80 and 250 Subjects. (MacQuarrie '27, p. 337)
Validity: r = .20 and .002 with group mental tests; .48 with teacher ratings of mechanical ability; and .32 and .81 with actual mechanical work. (MacQuarrie '27, p. 337)

r = .15 ± .13 for auto mechanics group; .25 ± .12 for the electrical group; .27 ± .14 for the machine shop group; .63 ± .10 for the mill cabinet group; with an average r of .30 ± .06 for all the groups. (Kefauver '29, p. 202)

Discussion: MacQuarrie feels that this test measures something different from the traits measured by group mental tests; he notes that speed is an important element in mechanical success. He reports that teachers were enthusiastic over the test and that it may be taken by girls as well as by boys with no handicap.

Kefauver found that this test ranked midway between the Stenquist and the Terman Group Tests in correlating with ranks in shop success in general. When the subject groups were considered separately, however, there was a large variation in the measuring power of these three tests.

References:

MacQuarrie '27 Kefauver '29

No. 60 Manual Dexterity Tests No. I and II

Devised by H. C. Link.
Date: 1919
Handled by Stoelting.
Price: $16.00 each number.
Material: Link ('19) reported two boards, No.1 having two sets of square insets, graduated in size; No. 2 having a similar series of "odd shapes and sizes." Stoelting lists a board made of bakelite, stating that the two original sections of Link's No. 1 board are now combined.
Problem: After the pieces are turned over in the cover of the board and left in their exact order, the Subject must put them back in their proper places. This is tried with each hand alternately, then with both together. Subject must work from the smallest piece to the largest.
Individual test; not adaptable to group testing.
Directions: Verbal and pantomime.
Responses: Manipulative.
Time: No limit.
Scoring: The hand used, and time, and errors.
Standardization: Link reports the results of using this test in industry, but detailed data as to standardization are not given. (Link '19, p. 55 ff)
Validity: r= .72 with ratings of "superior" and Test II (odd sizes and shapes). r= .52 with combined ratings of two judges for twelve girl assemblers on combined manual dexterity tests. (p. 59) r = .32, combined ratings of two judges for thirty-one men assemblers. r = .26, combined ratings of two judges for twenty-six men assemblers. r = .42, on thirty-one men with Stenquist Assembly Test. r = .58, on twenty-six men with Stenquist Assembly Test. (Link '19, p. 59 and 75)
Discussion: The tests have been found effective in selecting left-handed people. (Link '19, p. 429) The tests call for the ability to detect slight differences; to size up a space or a piece; to make a quick decision, and to match the piece with the space. Some note quickly where these parts belong, while others aimlessly try out each piece in one opening after another. (Link '19, p. 56 f).

References:

Link '19 Wallin '27

No. 61 Manual Dexterity Tests

Devised by E. C. Whitman and J. O'Connor.

Date: 1925

Handled by Stoelting.

Price: $11.50

Material: A metal tray divided into five sections; three blocks about 12 x 6 x 1 inches, with suitably drilled holes; 100 colored wooden pegs, 200 brass pins; and twenty bolts with nuts.

Problems: 1. To put pins into holes quickly, one at a time, using only one hand. 2. Repeat, using other hand. 3. To put three pins in each hole, using both hands. 4. With the right hand, to insert pegs into board according to a specified color and space arrangement. 5. After demonstration by Examiner, Subject is required to separate nuts from bolts. 6. To put the nuts and bolts together again. 7. With the right hand, to sort pegs into the four compartments according to color.

Individual test; not adaptable to group testing.

Directions: Verbal.

Responses: Manipulative.

Time: Parts 1 and 2: one minute each. Part 3: two minutes. Part 4: one minute. Parts 5, 6, 7: half minute each.

Scoring: Final score is the sum of credits for each item as set down by Whitman ('25).

Standardization: Whitman has standardized this test on 491 cases. Scores for each item separately and total scores are distributed in terms of age, and quartiles for each age. Range: Seven to fifteen years. (Whitman '25) Norms are also expressed for time, and arranged in quartiles for males and females. (O'Connor, letter to writers)

Discussion: This series is not an attempt to test intelligence, mechanical ingenuity or any special ability other than that implied in its title. Apparently, however, it will serve a need in the clinic as well as in industry. (Bronner, Healy *et al.* '27, p. 211)

References:
Whitman '25
Bronner, Healy, Lowe, Shimberg '27
O'Connor '28, and letter to writers.

No. 62 Measures of Musical Talent

Devised by C. E. Seashore; also modified by J. Kwalwasser.

Date: 1919

Handled by Stoelting; also by Columbia Phonograph and Victor Talking Machine Co.

Price: 0.40 (Stoelting). $1.25 each record (Columbia Phonograph); manual of instructions and interpretations free.

Material: A set of five double disc phonograph records.

Problem: Subject must judge as to the similarity or difference, or comparative beauty of a paired series of notes which include tests of: 1, pitch; 2, intensity; 3, time; 4, consonance; 5, memory; and 6, rhythm. One hundred trials for each series.

Individual or group tests.

Directions: Verbal.

Responses: Verbal.

Time: About 90 minutes

Scoring; Percent right.

Standardization: Three sets of percentile scores are given in the manual of directions for 1, adult; 2, grade VIII; and 3, grade V. Children of the sixth and seventh grades may be assigned proportional ranks in between.

Validity: Studies on the validity of the Seashore musical tests have been made by Brennen ('26); Stanton ('25) and Brown ('28). Brown summarizes the previous two studies, and his own results, are not inconsistent with them. His criterion was the average of two ratings, given four months apart, by the high school music teacher. These ratings were carefully checked for reliability. ($r = .94 \pm .01$)

Validity:

Pitch	$r = .15 \pm .06$	Time	$r = .15 \pm .06$
Intensity	$r = .11 \pm .06$	Consonance	$r = .17 \pm .06$
Memory	$r = .41 \pm .05$	Rhythm	$r = .17 \pm .06$

Average score $r = .38 \pm .05$ (Brown '28, p. 473)

Reliability: Seashore claims that the degree of reliability obtained by different investigators is dependent on their general experience and freedom from bias. He sums up results as follows: For adults, $r = .71$ (R. Larson): $r = .63$ (D. Larson); $r = .68$ (Peterson); $r = .57$ (Gaw); $r = .69$ (Salisbury and Smith) "all with satisfactory probable errors." Lower correlations are obtained by Brown ($r = .43$) and by Weaver ($r = .35$) and by Heinlein. In their present form the tests are not satisfactory for use with children.

Discussion: Seashore claims that the physiological limit of pitch discrimination does not vary with age, or with intelligence; that pitch discrimination does not vary with sex, that it is a capacity which cannot be appreciably improved by practice. In regard to the ability to discriminate for duration of time interval, age, training and intelligence have but slight effect. ($r = .17 \pm .04$) Rhythm is instinctive and shows very wide individual differences among people. The sense of consonance, too, is a gift which, like those of pitch and of rhythm, is revealed early in childhood. Musical memory can be rather extensively cultivated through training. (Seashore '19)

Brown's findings confirm Seashore's claims of little or no correspondence between age and musical ability, and intelligence and musical ability, as measured by these tests. There is also but little intercorrelation between the different elements which go to make up the composite "musical ability." Brown says, "The ability to respond to these isolated factors may be quite different from the ability to respond to a combination of them. In fact, a combination of the elements may present a situation which has very little relation to the elements of it. This fact has been well illustrated in the development of intelligence tests." (Brown '28, p. 474 f)

References:

C. E. Seashore, '19 and '30
Kwalwasser, '27 and '28
Schoen, '23
Brennan, '26
Lanier, '27
Farnsworth, '28
McGinnis, '28
Larson, '28
Brown, '28
Smith and Wright, '28 (p. 284f)
Stanton, 28 a and b
Broom, '29
Davenport, '29
Foster and Tinker, '29
Peterson and Lanier, '29

No. 63 Mechanical Aptitude Tests, I and II

Devised by J. L. Stenquist.
Date: 1923
Published by The World Book Company, and Stoelting.
Price: $1.25 per 25 for each set. (World Book Co.) Manual, pp. 21, $0.15. $2.50 per 25 for each set (Stoelting).
Material: Paper booklets, each of seven pages. Test I contains ninety-five pictures of mechanical objects, with questions about relationships. Test II is in part, similar to Test I.
Problem: Test 1: Subject must determine which of five pictures belongs with each of five other pictures. Test II: Subject must answer questions relating to a machine and its mechanical parts.
Individual or group test.
Directions: Verbal. Part I, pantomime adaptation possible; Part II, not.
Responses: Non-verbal in Part I.
Time: Test I, forty-five minutes; Test II, fifty minutes.
Scoring: Raw scores (number right) converted into T-score equivalents and percentile ranks for each age.
Standardization: This test was standardized by its author on 664 and 1087 cases for Test I and Test II respectively, in Grades VI and through VIII and high school; ages, ten through fifteen. T-score and percentile distributions are supplied.
Reliability: For test I, r = .79 on 103 cases in Grades VI, VII, VIII. For Test II, r = .61 (split half) between Exercises 1 and 2, and r = .68 between Exercises 2 and 3, on 200 unselected cases in Grades VI, VII, VIII. (Stenquist '23, p. 73)
Validity: r's range from .44 to .88 against a criterion of Stenquist Mechanical Assembly Tests. r's range from .43 to .83 against shop teachers' ranks. (Stenquist '23, p. 75)
Discussion: From the evidence of the above correlations, Stenquist adds: "It is therefore entirely justifiable to assume in general that a high score in the picture tests is an indication of general mechanical aptitude. To obtain the best measure, both the assembly tests and the picture tests are advisable. For preliminary classification, however, the picture test alone may serve." (Stenquist '23, p. 75)
References:

Stenquist '22 and '23
Toops '23
Kefauver '29

No. 64 Metal Puzzles. Series I

Devised by H. A. Ruger; modified by F. W. Steacy.
Date: 1919.
Material: Each puzzle consists of two parts. These in general are made of different grades of wire, pieces of metal, pieces of cord and so forth. The eight puzzles are contained in a stout cardboard box, 13½ x 6 x 1¼ inches. The box has several compartments in which puzzles are arranged according to estimated difficulty. (Steacy '19, p. 9)
Problem: To separate each puzzle into two parts without the use of force.
Individual test; eight different puzzles. This test is adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Fifteen minutes.
Scoring: One mark for each puzzle separated without use of force.
Standardization: Steacy experimented with these puzzles on thirty-one boys and thirty-eight girls in the sixth grade of a Bronx public school. Average deviations are supplied for this group. Otherwise, the series is not adequately standardized.
Reliability: r = .32. The smallness of this correlation is due probably to the element of chance which enters so largely into all kinds of puzzle tests. (Steacy '19, p. 40)
Validity: r = .45 against Steacy's battery of tests, excluding handwriting. (Steacy '19, p. 40)
Discussion: This test, according to Steacy, requires analytic ability. "The simpler puzzles may be solved by merely fumbling with them, but the more complex puzzles require some insight into their construction. The Subject does not need any knowledge of patterns or of mechanical instruments. He can clearly see the whole of each component part. He needs to know only the method by which the parts can be separated." (Steacy '19, p. 39).

F. L. Wells suggests that the Ruger puzzle tests might be used with advantage to study learning ability. ('27, p. 137)

References:
Poffenberger, '16
Steacy, '19
Yerkes (Ed.) '21
Perrin and Klein, '26
F. L. Wells, '27
Dashiell, '28
Turner and Betts, '24

N. B.: Ruger warns against using store puzzles since they are not manufactured under standard conditions, and are not "fool-proof," can often be solved in more ways than one, become known to the general public, and appear and disappear on the market. Obviously standardized manufacture, control and distribution are needed. (personal interview)

No. 65 Metal Puzzles. Series II

Devised by H. A. Ruger and F. W. Steacy.
Date: 1919
Material: Each puzzle consists of two parts; one part is a leather band, one-

quarter inch wide; the other is made of Number 10 wire. The three puzzles are contained in a stout cardboard box, 7 x 5½ x 1 inches. Each box has three compartments, one for each puzzle. (Steacy '19, p. 10)

Problem: To remove the leather strip from the metal part without the use of force.

Individual test; three puzzles increasing in difficulty. This test is adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Fifteen minutes.

Scoring: A mark is given for each correct movement. The first puzzle has four movements; the second has nine; the third has nineteen. As the puzzles are usually returned wholly apart or wholly together, very few marks are given for partial solutions.

Standardization: This series is not standardized, but Steacy experimented with these puzzles on thirty-one boys and thirty-eight girls in a sixth grade New York public school group. Average deviations are stated.

For reliability and validity: See Series I, Test No. 64

Discussion: While recognizing the general unreliability of puzzles as tests, nevertheless Steacy points out their value in measuring to some extent the Subject's learning ability. There is a common underlying principle in the three graduated series which, if discovered, will make the solution of succeeding puzzles much easier than it otherwise would be, for the increasing complexity of the puzzles involves increasing difficulty for the solver. Steacy '19)

References:

Steacy, '19

N. B.: See foot-note to Metal Puzzles, Series I. No. 64.

No. 66 Metal Puzzles. Series III

The basis consists of Chinese rings which can be purchased ready-made. These puzzles were devised by H. A. Ruger and modified by F. W. Steacy to form Series III.

Date: 1919

Material: Five puzzles, all identical except in the number of rings which vary from one to five and which are protected by a piece of strong string and sealing wax. The five puzzles are contained in a box with five compartments, 13½ x 6 x 1¼ inches. The puzzles are arranged according to difficulty. A paper clip is placed over the point of the bar to prevent slipping off. (Steacy '19, p. 10)

Problem: To separate the frame, carrying the rings, from the bar without damaging the puzzles.

Individual test; five problems. This test is adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Thirty minutes.

Scoring: One mark is allowed for each movement properly made, comprising fifty-nine in all.

Standardization: This test is not standardized, but was tried out experimentally by Steacy on thirty-one boys and thirty-eight girls of the sixth grade in a New York public school. Average deviations are given.

For reliability and validity: See series I, Test No. 64.

References:

Steacy '19

N. B.: See foot-note to Metal Puzzles, Series I, No. 64.

No. 67 Metal Puzzles. Series IV

The basis of this material can be purchased ready made; it was devised by H. A. Ruger and modified by F. L. Steacy to form Series IV.

Date: 1919

Material: Boxes, $12\frac{1}{2} \times 7\frac{3}{4} \times 1$ inches, each box having fifteen compartments, one for each puzzle. All these puzzles are made of stout steel wire, nickel plated, except one, which is made of cast iron. (Steacy '19, p. 11)

Problem: To separate the puzzles into their parts.

Individual test; fifteen puzzles, no two being alike, although three form one group, differing only in the number and arrangements of the parts, while four others form another group, likewise differing only in the number or the arrangements of the parts. This series is adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Thirty minutes.

Scoring: Partial credit allowed according to the number of correct moves. Total possible is forty-eight marks.

Standardization: This material has not been standardized, although Steacy tried it out experimentally on thirty-one boys and thirty-eight girls in the sixth grade of a New York City public school.

Reference:

Steacy '19

N.B. See foot-note to Metal Puzzles, Series I, No. 64

No. 68 Minnesota Assembly Test

Devised by J. L. Stenquist; modified and amplified by L. D. Anderson, *et al.*

Date: 1929

Handled by Marietta; also by Stoelting.

Price: Long Form, $29.00, (Marietta); $32.25, (Stoelting). Short Form, $17.00, (Marietta); $19.00, (Stoelting). Score Cards, $1.00 per hundred (Marietta).

Material: Long form: Three boxes, containing various disassembled mechanical devices,—twenty-four, plus Stenquist's original ten, of which two were changed. There are also two short forms, Sets I and II of equivalent difficulty, each containing material to make up ten objects.

Box A	Box B	Box C
1. Expansion nut.	1. Safety razor.	1. Harmostat.
2. Hose pinch clamp.	2. Monkey wrench.	2. Die holder.
3. Hunt paper clip.	3. Ringstand clamp.	3. Pliers.
4. Wooden pinch clothes pin.	4. Test-tube holder.	4. Electric light socket
5. Links of chain (6).	5. Spark plug.	5. Wing nut.
6. Bottle stopper.	6. Inside callipers.	6. Glass drawer knob.
7. Push button door-bell.	7. Electric plug and wire.	7. Rope coupling.
8. Bicycle bell.	8. Clover-leaf coin purse.	8. Kettle cover knob.
9. Plug and wire.	9. Flat-iron handle	9. Lock nut.
10. Corbin rim lock	10. Mouse trap.	10. Fork Magneto-post.
		11. Petcock.
		12. Hose clamp.
		13. Radio switch.
		14. Pencil sharpener.
		15. Air gauge valve.
		16. Mechanical pencil.

Problem: To assemble the parts of each device.

Individual or group test. If the latter, use blackboard diagrams as set down in (Manual p. 8)

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Time limits set for each item. About an hour's working time is required to do the long form. (Manual p. 10)

Scoring: Number of parts correctly assembled; weighted on the basis of ten; to be recorded on score cards. (See Manual p. 15 ff)

Standardization: This test has been very carefully standardized by workers at the University of Minnesota. Raw scores are convertible into percentile rank in a given age, sex, or academic grade group. They are stated separately for the long form, for the short form and for each set. Range: ages, from eleven to above twenty-one; grade VII through university group. (Manual in mimeograph)

Reliability: r= .94 (odd-evens coefficient, treated by Brown-Spearman prophecy formula). (Anderson '28b)

Validity: A high validity was obtained for these tests by weighting the scores of the individual tests and adding the results. This resulted in a battery which was much more highly predictive of mechanical ability than was any single test in the group. Using a criterion of scores in shop success (See our No. 70) 1, fifty per cent of the cases were located exactly, e. g. boys with test score A received shop rankings A. 2, forty-five per cent were located one position away. 3, five per cent were located two positions away. (Anderson '28b)

Discussion: Because of the very definite success of these tests in predicting capacity for success in mechanical shop courses, it is important that their value be determined in actual industrial situations as well. Anderson '28b, p. 478)

References:
- H. W. Rogers, '24
- L. D. Anderson, '26, '27, '28a, '28b
- Edgerton and Paterson, '26
- Paterson, '28
- Hubbard , '28
- Carter, '28
- Peterson and Lanier, '29
- 'Proceedings', (Riley) '30
- Paterson, Elliott, Anderson, Toops and Heidbreder. To be published by the University of Minnesota Press.

No. 69 Needle-Threading Test

Devised by J. H. Hayes and E. Dewey.
Date: 1920
Handled by Stoelting.
Price: $4.00
Material: Set of five steel needles, sizes 7, 8, 9, 10, set in a wooden block. One end of each needle is flattened and has a hole drilled into it. Spool of cotton thread.
Problem: To thread the needles, beginning with the largest.
Individual test.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Limit of two minutes on each needle.
Scoring: Number of needles threaded. Repeat until subject fails two successive needles. ('20, p. 35)
Standardization: This test was used by Dewey, Child and Ruml. These norms are offered separately for each sex, for ages nine to thirteen inclusive, 100 children to each age. Means and S. D's are stated. ('20, p. 80)
Discussion: Apparently little use has been made of this test. Tables prepared by Dewey, Child and Ruml show a slight and not always consistent age-to-age increase. The test was retained by them for inclusion in their "maturity scale." ('20, p. 80)
References:
- Dewey, Child, Ruml, '20

No. 70 Objective Measurements in Shop Courses

Devised by L. D. Anderson *et al.*
Date: 1926
Materials consist of three measuring instruments and four rating scales against which shop work is rated.
Instruments are: (1) dimension meter; (2) squareness machine; (3) wind gauge. The rating scales were constructed for:

1. Electricity: tap splice; Western Union splice; pig tail splice.
2. Woodwork: boring, sawing, 3 scales.
3. Sheet metal: soldering, 6 scales; riveting.
4. Mechanical drawing: lettering, 3 scales; numbering; fractions; blackness of lines.
5. Printing: several factors.

Problems: The manipulative ability of pupils in shop work is graded by the above measuring devices. The scales alone can make a total of 1314 separate measurements on each boy's projects in sheet metal, woodwork, electricity and mechanical drawing, based on job analyses of work done in the shop course.

Directions: Since this is not a "test," but a measuring scale, no strict task is set. A varying number of projects which require the use of tools may be rated.

Responses: Manipulative.

Time: Two hours are required to grade a set of twenty woodwork or sheet metal projects.

Scoring: Objective reading on measuring device; place location on rating scales.

Standardization: To construct the scales, one hundred or more samples of work done by junior high school boys, were rated by six to twelve judges, and ranked for quality. From these, twenty samples were selected for scale points.

Reliability: $r = .76$. (Anderson '28, p. 147)

Discussion: "This brief description of the measuring devices and tests developed in connection with the study on mechanical abilities indicates the practicability of using these or similar methods in the actual grading of projects in the courses when the results are not to be used for research purposes." (Anderson '26, p. 267)

References:

L. D. Anderson, '26 and '28

No. 71 Puzzle Box

Devised by F. N. Freeman.

Date: 1916

Handled by Stoelting; also by Marietta.

Price: \$62.50 (Stoelting); \$45.00 (Marietta).

Material: Puzzle box with a system of levers which will open the box.

Problem: To open the box without using force.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Five minutes (Woolley). "About ten minutes, but upper limit indefinite." (Freeman, letter to writers).

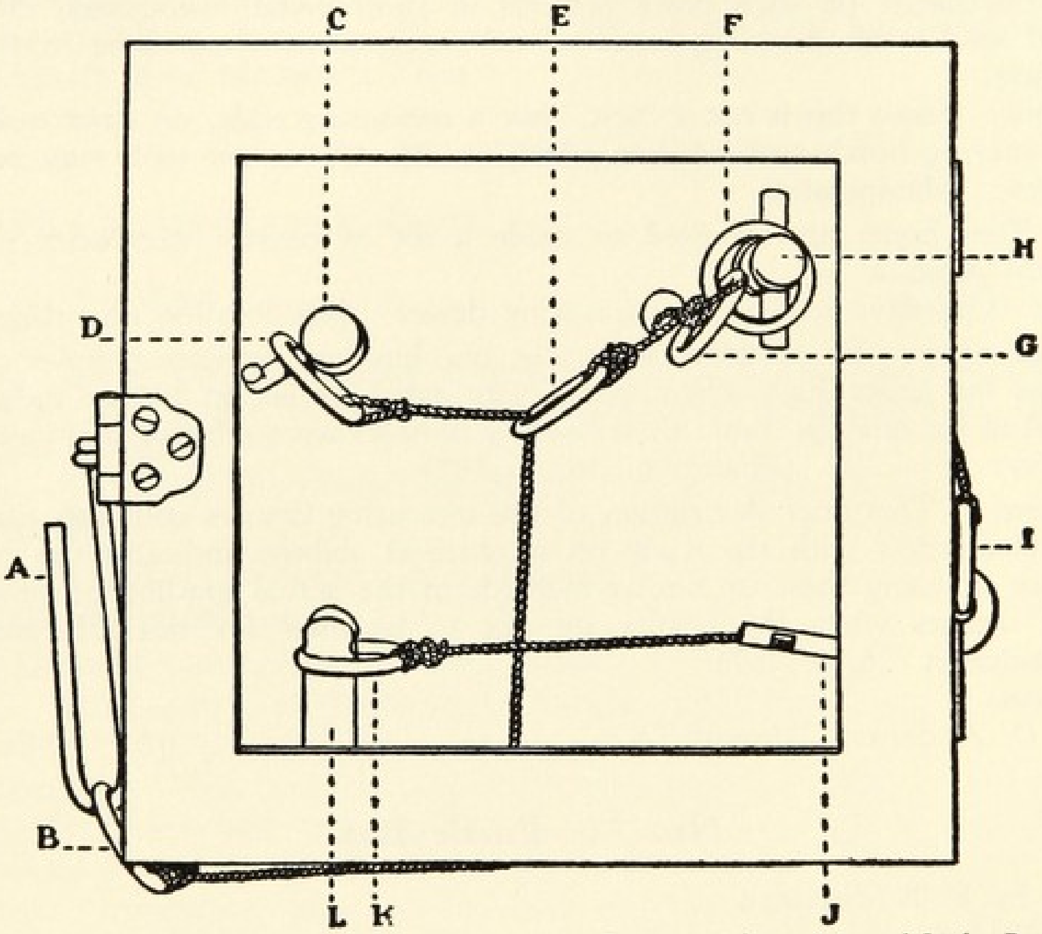
Scoring: Time.

Standardization: Freeman reports that this test is not yet adequately standardized, but that it may be in a few months. Woolley tried it on 336 males and 232 females, both groups being eighteen-years-olds. She distributed the scores (time) by percentiles and sex. Freeman thinks that the test will be suitable for ages from about twelve years up. (letter to writers)

References:

F. N. Freeman, '16 (and letter to writers)
Averill, '24
Woolley, '26
Bronner, Healy, Lowe and Shimberg, '27

No. 72 Puzzle Box



Reproduced by courtesy of Little, Brown & Co.

Devised by W. Healy, M. Fernald and J. Hayes.
Date: 1911
Handled by Stoelting.
Price: $26.50
Material: Puzzle box with glazed top which enables the Subject to observe the movements necessary to open the box, which is kept closed by a system of string attachments. These can be unfastened by a buttonhook inserted through holes in the side of the box. A modification of this test by Woolley is painted black inside, while this is light colored.
Problems: By manipulation of levers, to open the box; seven steps.
Individuals: Verbal; pantomime with difficulty.
Responses: Manipulative.
Time: Fifteen minutes: five minutes to open; ten minutes to close.

Scoring: Note method of attack and procedure; time.

Standardization: This test was standardized by Shimberg, Lowe and Meehan, on 277 cases. Norms are expressed in terms of sex and in percentiles on a group, aged fourteen and one-half years through adult.

Validity: McFarlane found a positive correlation between the Puzzle Box and Cube Construction in sixty-six per cent of a group of technically trained children and none in a purely random group.

Discussion: Healy felt that the test brought out "manipulative powers, ability to analyse a slightly complicated physical situation, in powers of attention, and continuity of effort." (Healy and Fernald, '11, p. 18). McFarlane considers it a test of "practical ability" as expressed in "the power to grasp directly spatial, temporal and qualitative relations between concrete things." (McFarlane, '24, p. 55f) The test "requires good judgment, based upon visual and kinaesthetic clues, and ability to keep the successive steps in mind long enough to indicate the next one." (McFarlane, '24, p. 37)

References:

Healy and Fernald, '11
Woolley, '14, '26
Healy, '15
Averill, '24
McFarlane, '24
Bronner, Healy, Lowe, Shimberg, '27

VI

CONSTRUCTIVE ABILITY TESTS

No. 73 Automobile Construction Test

(Cart Construction Test)

Devised by J. W. Hayes; modified by E. Dewey.
Date: 1920
Handled by Stoelting.
Price: $1.75

Figure 1

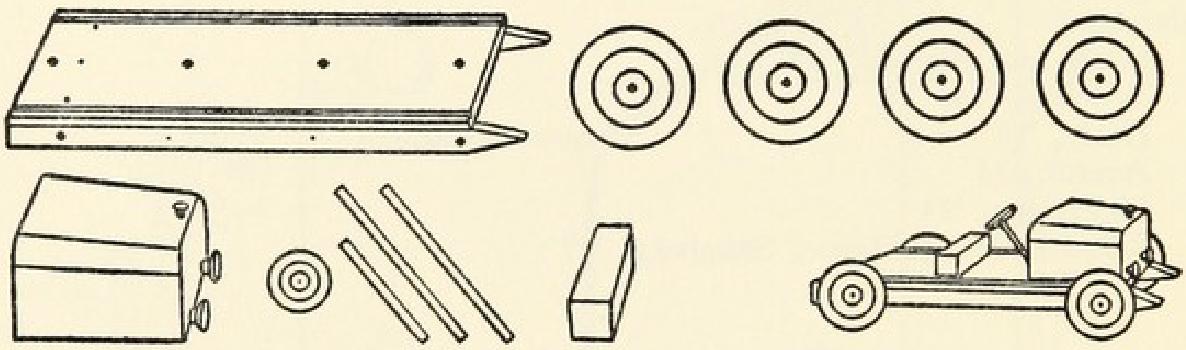
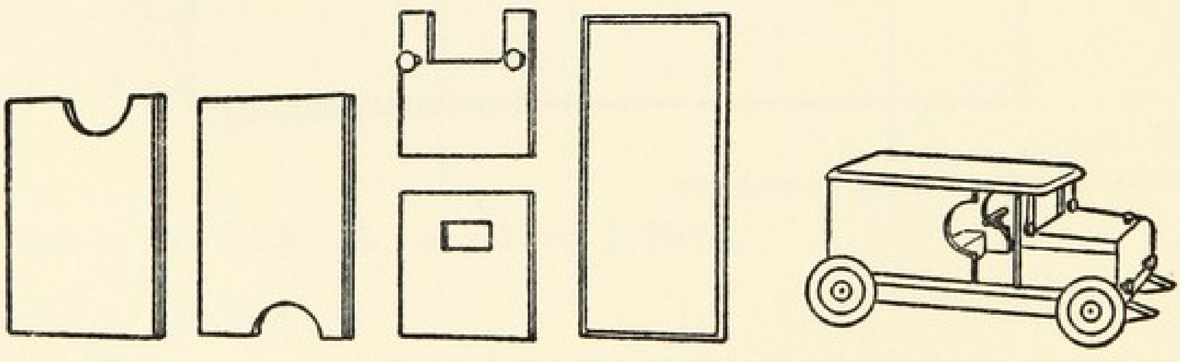


Figure 2



Reproduced by courtesy of Little, Brown & Co.

Material: The set consists of sixteen parts of a wooden toy automobile. (Schoenhut's Five-in-One Auto Build).
Problem: (1) To construct an automobile from its parts; (2) to construct a delivery wagon.
Individual test; two parts; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Eight minutes for each part.

Scoring: Move by move according to a table of points. (Dewey *et al*, p. 22)

Standardization: This test was standardized by Dewey, Child and Ruml, on 500 Jewish children. Norms are expressed in terms of means for each age and sex, and separately for each part. The test is adaptable to children from nine to thirteen years. (Dewey *et al*, p. 78)

Discussion: This is a test of constructive ability, but also dependent to a large extent on local knowledge. Dewey *et al* felt that, because of the simplicity of its material, the test "is one involving definite constructive or creative ability, rather than mechanical ability." (p. 23)

References:

Dewey, Child and Ruml, '20
Bronner, Healy, Lowe and Shimberg, '27

No. 74 Bead-Stringing Test

Devised by H. K. Wolfe.

Date: 1927

Material: Two long spindles, one empty, the other strung with twenty-six beads, to be used as a sample string. A pan of Mrs. Hailmann's kindergarten beads, obtained from Milton, Bradley. These come in six colors: red, orange, yellow, green, blue, purple; and in three different shapes: sphere, cylinder, cube. The beads on the sample string are so arranged that no two of the same color or same shape are adjacent. This material is carried in a wooden case, 17¾ x 3⅝ x 3⅝ inches.

Problem: To string the beads on the empty wire, similar to the model.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative. This test could be adapted to small groups.

Time: No time limit.

Scoring: Time plus errors. Errors are counted as 1-26th of total time.

Standardization: This test was standardized by Blacking for raw scores, time and errors in terms of age on 422 unselected American and foreign born children. Range: six to sixteen years.

Validity:

r= .23 against Otis test of mental ability for 36 fifth graders.
r= .257 against Illinois Intelligence Test for 70 sixth and seventh graders.
r= .460 against Dearborn Group Test for 45 cases, grades one to three.
r = .532 against Detroit First Grade Intelligence Test, 47 cases, grades one and two.
r= .096 against McCall Multi-Mental Scale for 32 sixth graders. (p. 632)

Discussion: "Coefficients of correlation with other tests are high enough to show that the test is a measure of some phase of intelligence." (Blacking, '27, p. 633)

Reference:

Blacking, '27

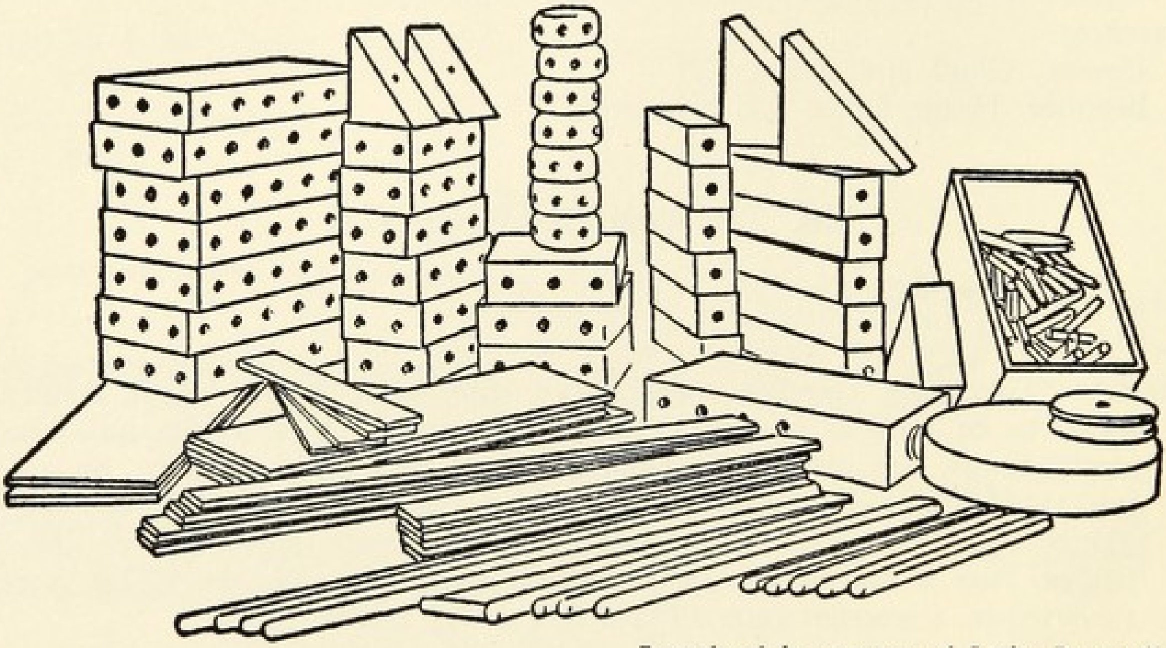
No. 75 Construction Ability Test

Devised by T. L. Kelley.

Date: 1916

Handled by Stoelting.

Price: $15.00; and $0.80 per hundred for record sheets; also $15.00 for set of thirty-nine stereoscopic pictures to be used in scoring the test. This includes a copy of Kelley's article.



Reproduced by courtesy of Little, Brown & Co.

Material: Set of drilled blocks, boards, dowel pins, and so forth; the bulk of a small typewriter case.

Problem: To construct anything the Subject wishes from the material supplied.

Directions: Verbal; pantomime adaptation doubtful.

Responses: Manipulative.

Time: Twice the number of minutes as Subject's age in years.

Scoring: By comparison with stereoscopic photographs which constitute a scale.

Standardization: This test has never been adequately standardized. The author considers it suitable for persons from one and one-half years to adult, for all groups. (letter to writers)

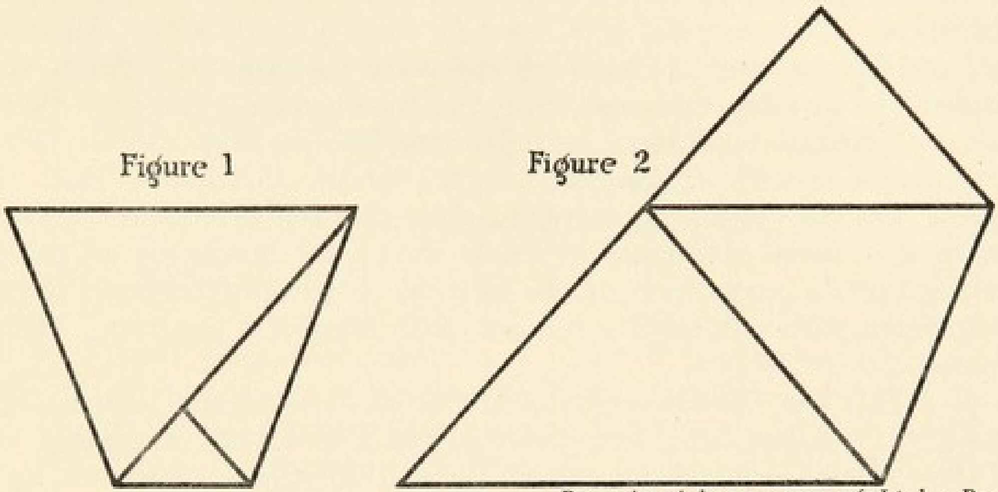
Discussion: This test, according to its author, measures "originality and initiative in mechanical manipulation." (letter) Pintner and Paterson criticize it as "being a toy in very common use among children. Some children will have opportunity to be more familiar with this than others." (Pintner and Paterson, '17, p. 22)

References:

Kelley, '16, and letter to writers
Pintner and Paterson, '17

No. 76 Folded Drinking Cup Test

Devised by the New York Bureau of Analysis and Investigation.
Date: 1915



Reproduced by courtesy of Little, Brown & Co.

Materials: Two pieces of typewriting paper, 8½ x 8½ inches.
Problem: After a demonstration has been given, Subject is required to construct a cup by folding the paper.
Individual test; not adaptable to group testing with the present method of scoring.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: From sixty-seven to forty-three seconds.
Scoring: Time and procedure.
Standardization: This test was reported and standardized by the New York Bureau of Analysis and Investigation on children from seven to fifteen years. There is but slight variation in time or number of moves for the different ages.
Discussion: This test is too simple above ten years, "but below that level it affords a useful way, without special equipment, of testing ability to follow easy directions. The test is often invalidated, however, by the subject's previous knowledge of how to make such a cup." (Bronner, Healy, *et al*, '27, p. 199)
References:
New York Bureau of Analysis and Investigation, '15
Bronner, Healy, Lowe, Shimberg '27

No. 77 Form and Assembling Test

Devised by S. D. Porteus.
Date: 1926
Material: Twenty-two cards with pictures on them of various parts of a knife, hammer, wheelbarrow, coffee pot and chair.
Problem: First part: to match as quickly as possible the pictures of the above common objects. Second part: to select as quickly as possible the two parts that make the wheelbarrow, the three parts that make the knife, the four parts that make the coffee pot, and the eleven parts that make the chair.

Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring: Total time for all the operations in seconds, taken with a stop watch.
Standardization: This test has been used on Hawaiians, Chinese, Japanese, Americans and Portuguese. Norms are expressed in terms of seconds required to complete the test for each age from five to nineteen. Range: Best for ages seven to fourteen. (Porteus and Babcock '26, p. 239)
Validity: r = .33 with Thorndike High School Graduate Test. However, the test was not found to correlate with university students' college marks. "There is a form of mental alertness shown by quickness of perception of form and ready recognition of the relation of parts of common objects which is associated with test solving but not with scholastic progress." (Porteus and Babcock '26, p. 242)
Discussion: This test "combines the features of a formboard and a logical relations test". It has been found to correlate well with industrial and social abilities. "It is primarily a test of mental alertness, as shown in speed of recognition of forms either singly or in combination with other forms." (Porteus and Babcock, '26, p. 238)
Reference:
Porteus and Babcock, '26

No. 78 Furniture Test

Devised by W. F. Dearborn and A. O. Christiansen; modified by E. A. Shaw.
Date: 1916
Handled by Psycho-Educational Clinic, Palfrey House, Harvard; also by Stoelting.
Price: $13.50 (Stoelting)
Material: Fifteen pieces of light weight wood, mostly in "stick form".
Problem: To make a piece of furniture using all the material; three trials.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime adaptation doubtful.
Responses: Manipulative.
Time: Fifteen minutes or less.
Scoring: Time.
Number of pieces used, one point for each.
Final score: best score for each trial.
Standardization: This test is not adequately standardized. Scores are reported however, for ninety children, aged seven to thirteen years. (Dearborn *et al*, '23, p. 62 f)
Discussion: Dearborn considers that tests of this sort are more general in their nature than are form-boards and that they require abstract thinking in some degree for their successful completion. Referring to furniture tests, he says: "It is possible that we shall discover that performance tests of general intelligence must be of this latter type." (Dearborn, Shaw and Lincoln, '23, p. 64) Healy, Bronner, *et al*, remark that in a test of this

sort, "important features of the temperamental and emotional make-up of the child may often be observed by his approach to the problem and his method of solution." They also draw attention to the familiarity of children nowadays to this sort of material in their toys and games. ('27, p. 210)

References:

Dearborn, Anderson and Christiansen '16
Dearborn, Shaw and Lincoln, '23
Bronner, Healy, Lowe, Shimberg, '27

No. 79 Intelligence by Drawing

Devised by F. L. Goodenough.

Date: 1926

Published by World Book Company.

Price: $1.00 for twenty-five test sheets; $1.80 for manual.

Material: Single sheet of paper and a pencil.

Problem: To "draw the picture of a man".

Individual or group test.

Directions: Verbal: pantomime by adaptation; or could be translated without modifying essential conditions. (Goodenough, letter)

Responses: Non-verbal.

Time: About ten minutes usually; no formal time limit.

Scoring: Single performance scored from four different standpoints. Points according to directions in manual, or in code on each test sheet.

Standardization: This test has been carefully standardized by Goodenough on 3593 primary grade children. Norms are expressed in terms of raw scores, translated into mental age or percentiles. The test is adaptable to children from five to ten years.

Reliability: r = .937 (test-retest method) for first grade children.
r = .77 (split scale method) averages for ages five to ten taken separately.

Validity: r = .74. Drawing I. Q. vs. Stanford-Binet I. Q. S. D. for Drawing is 21.2; for Stanford-Binet is 19.1. (Goodenough, letter to writers)

Discussion: Goodenough ('26c) found that the test showed definite differences between groups of Negroes, Indians and various nationalities. This tended to conform to results from other tests. Blackwood suggests that this sort of test is a good one to use with primitive peoples, and points out a common experience in marking and drawing. She found from her experience in giving a similar test that young children who had had no schooling, did comparatively well. ('27, p. 54 f) Goodenough reports that this test is not a measurement of ability in drawing, but of general intelligence. For opinions as to how much this test is free from local knowledge, see pp. 53-55 of Goodenough's manual. Prof. Ed. Claparede of Geneva is now testing it for this purpose.

References:

Goodenough, '26a, b, and c, and letter to writers
Blackwood, 27
Yepsen, '29
Davenport, '29

No. 80 Tower Test (Nest of Cubes Test)

Devised by New York Bureau of Analysis and Investigation.
Date: 1917
Handled by toy stores; also by Stoelting.
Price: $1.70 (Stoelting)
Material: Nest of seven square boxes, graduated in size from 3½ to 2½ inches, with one side open.
Problem:

1. To build the tallest tower possible out of the blocks.
2. To raise and remove the tower to a table.
3. To pack the blocks away.

Individual test; three parts. The test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Five minutes or less.
Scoring: Time and success on each of the three parts.
Standardization: This test was reported and standardized by the New York Bureau of Analysis and Investigation on 728 institutional children. Norms are expressed in terms of time for building, raising, removing and packing the blocks. Range: for mental ages from six to twelve years.
Discussion: The Experimenter can observe the Subject, "from the standpoint of his ability to interpret and follow directions, his precision and mode of attack on a rather delicate task and his determination to succeed." (New York Bureau '17, p. 130)
References:

N. Y. Bureau of Analysis and Investigation, '17
Carlisle '18
Woolley and Cleveland, '23
Stutsman, '26
Bronner, Healy, Lowe and Shimberg, '27

No. 81 Tying a Bow-knot

Devised by E. B. Huey; adapted by L. T. Terman.
Date: 1916
Published by Houghton-Miflin in conjunction with the Stanford-Binet tests.
Material: A model of a bow-knot, prepared in advance by tying a shoe-string around a small stick; an extra piece of string of the same length.
Problem: With the model in front of him, Subject must tie a bow-knot with the extra piece of string.
Individual test; adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: One minute.
Scoring: Test is passed if double bow-knot is made within one minute. Half credit allowed if single bow-knot only is made.

Standardization: This test was placed by Terman in the Stanford-Binet scale at the seven-year level, where 69 percent of the children succeeded in passing it. (Terman '16, p. 196 f)

Discussion: Terman considers that this simple test fulfills reasonably the requirements of a good test. He found that age, apart from mental age, had but little effect on achievement thus tending to support its inclusion as an intelligence test, a product of general rather than special training. He thinks that the chief determinants of success on this test are: "(1) Interest in common objective things; (2) ability to form permanent associative connections between successive motor coordinations (memory for a series of acts); and (3) skill in the acquisition of voluntary motor control. The last factor is probably much less important than the other two." Motor awkwardness, although it is apt to prolong the time needed to tie the knot, is not often a cause of a failure. (Terman '16, p. 198 f)

Reference:

Terman '16

VII

PICTURE TESTS

No. 82 Eight-Piece Picture

Devised by E. A. Doll.
Date: 1921
Material: A picture postcard and the eight pieces of a similar card, made by cutting the second card in half horizontally, and in quarters vertically. Or any suitable art classic type of picture, which meets the psychological requirements without constituting a puzzle, may be used.
Problem: To fit the parts together in their right order as quickly as possible.
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring: Time.
Standardization: This test was referred to in the Army memoirs, (Yerkes, Ed '21, p. 365) and reported as having been tried out by Doll on 185 inmates of Jeffersonville Reformatory. It was apparently discarded by the Army psychologists. Later, it was reported by Doll ('26), who offered data on a provisional standardization, based on the Stanford-Binet mental ages of the 185 adult inmates of Jeffersonville. These norms are stated in terms of time and mental age, and "point scores". The mental age range is from eight to twelve. Percentiles are offered for grouped ages; eight to nine, ten to twelve, thirteen to sixteen.
Reliability: $r = .68$ with the Twelve-Piece Picture. Test No. 83.
Validity: $r = .43$ against criterion of Stanford-Binet mental test.
$r = .53$ against the Two-Inch Cube Test.
Discussion: See Twelve-Piece Picture. Test No. 83.
References:
Yerkes (Ed) '21 (p. 365)
Doll, '26

No. 83 Twelve Piece Picture

Devised by E. A. Doll.
Date: 1921
Handled by no one at present.
Material: A picture postcard view, "In Old Kentucky," showing a cabin in the mountains with several Negroes standing about; twelve pieces of a similar card, cut into thirds horizontally and quarters vertically. If a substitute card is used, it must have the same considerations as set down by Doll.
Problem: To fit the pieces correctly together as quickly as possible.

Individual or group test.
Directions: Verbal or pantomime.
Responses: Manipulative.
Scoring: Time.
Standardization: This test was referred to in the Army memoirs and later reported by Doll as having been experimentally tried out on a group of 185 inmates at Jeffersonville Reformatory. It apparently was discarded by the Army staff. (Yerkes, Ed., '21, p. 365) The provisional standardization offered by Doll is in terms of time, standardized against Stanford-Binet mental ages, and the norms are distributed for percentiles in groups: eight to nine; ten to twelve; thirteen to sixteen. "Point scores" are also offered.
Reliability: $r = .68$ with Eight-Piece Picture, Test No. 82.
Validity: $r = .49$ with Stanford-Binet.
$r = .55$ with Two-inch Painted Cube.
Discussion: Doll considers these tests (Twelve-Piece Picture and Eight-Piece Picture), although crude as to apparatus and procedure, as having definite suggestive value in the development of test materials for measuring the intelligence of illiterates and foreigners, because they meet the psychological requirements demanded by tests of this sort. ('26, p. 129)
References:
Yerkes, (Ed.), '21
Doll, '26

No. 84 Graded Series of Colored Picture Puzzles

Devised by G. H. Kent
Date: 1916
Material: Twenty pictures, mounted on cardboard, with corresponding pictures cut in pieces.
Problem: To reconstruct the picture, presented for a few seconds.
Discussion: Kent states that this test has no validity at all. The chief reason for publishing it was that the picture could be obtained at a low price, thus putting the materials within reach of any one who cared to make them. This no longer holds, as these pictures are not now on the market. The author wishes the series to be considered obsolete. (letter to writers)
References:
Kent '16a, and letter to writers.

No. 85 Mare and Foal Test

Devised by W. Healy and G. Fernald; modified by Pintner and Paterson.
Date: 1911
Handled by Stoelting; also by Marietta.
Price: $4.50
Material: Wooden board, 29 x 24.5 x 1 centimeters, on which is a picture of a mare and a foal; seven cut-out pieces. Pintner and Paterson modified the original test by sealing up the three geometrical forms which Healy and Fernald had cut out of the background. Maxfield has modified this latter

board by omitting the picture scene entirely but retaining the Pintner-Paterson cut-outs.

Problem: To put the pieces into their right places quickly.

Individual test; not adaptable to group testing.

Directions: Verbal or pantomime.

Responses: Manipulative.

Time limit: Five minutes.

Scoring: Time and errors.

Standardization: This test was revised and standardized by Pintner and Paterson on 621 school children. Norms are expressed for time and errors, in terms of age, percentiles and quartiles. Range: five through thirteen years. The test is especially applicable to children under eleven.

Reliability: "The distribution showed relatively little scattering." (Pintner and Paterson '17) Maxfield reports that the Pintner-Paterson norms seem to be a year too high. He has also found that when the picture is removed, the resulting distribution of scores remains the same; that is, that the picture form and the pictureless form are of equal difficulty. His results are based on a study of 476 children, aged five to nine years.

Discussion: Except for very young children, Healy believes the only value of this test lies in its attractiveness for the child, thus paving the way for further testing. (Bronner, Healy, Lowe and Shimberg, '27, p. 222) Maxfield has found that children pay but little attention to the picture itself, but rather to the form of the cut-outs. The picture, if anything, confuses them. He maintains that this test is a form board and not a picture test, and stresses the point that we should give a good deal of attention to the question of what tests test. (Reported at 1930 annual meeting, N. Y. Chapter, American Psychological Association)

References:

Healy and Fernald '11
Healy '15
Schmitt '15
Pintner and Paterson '17
Baldwin and Stecher '25
Stutsman '26
Bronner, Healy, Lowe, Shimberg, '27
Arthur, '28
Maxfield, personal interview

No. 86 Pictorial Completion Test I

Devised by W. Healy

Date: 1914

Handled by Stoelting; also by Marietta.

Price: \$13.25 (Stoelting); \$10.00 (Marietta)

Material: Board, 10 x 14 inches, on which is pasted a colored picture of a playground scene; fifty small one-inch square pieces.

Problem: To fit the most suitable pieces into the picture.

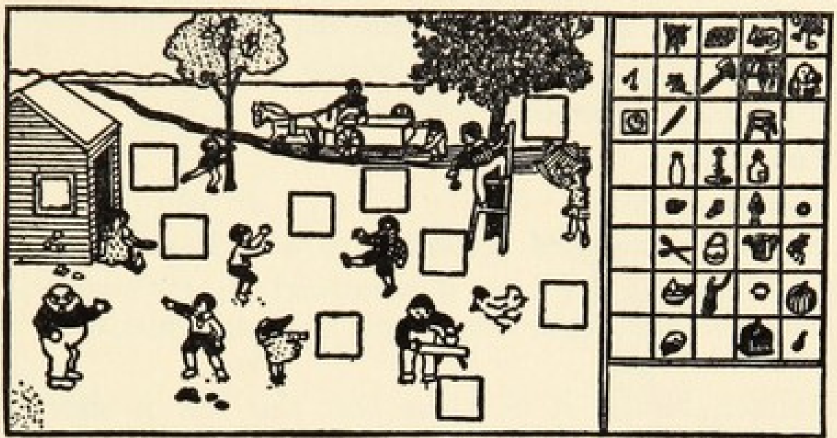
Individual test; not adaptable to group testing.

Directions: Verbal or pantomime.

Time limit: Ten minutes.

Scoring: Scores are based on credit given for each piece placed according to a table prepared by Pintner and Anderson ('17)

Standardization: This test was standardized by Pintner and Anderson on 1538 cases. Norms are expressed as raw scores and are distributed for age and percentiles. Range: five to sixteen years. Pictorial Completion I. does not discriminate among adults. (Pintner and Anderson '17, Healy '14)



Reproduced by courtesy of Little, Brown & Co.

Discussion: The Pictorial Completion Tests were offered originally by their maker as tests of intelligence in handling concrete material. However, Bronner, Healy, Lowe and Shimberg conclude from their experience with pictorial Completion Test I, that it is very uncertain in its significance because of the interfering factors of the cartoon-like drawing or humorous elements. They have found also that the norms set up by Pintner and Anderson ('17) run too low. For these reasons they have almost entirely discontinued using it in their clinic in favor of the later model, Pictorial Completion Test II. ('27, p. 185)

The introspections of seven psychologists who took these two tests, as reported by Gaw, indicate that the test calls for reasoning ability, that is, for the power to analyze and synthesize. ('25)

F. L. Wells considers that this test serves well for what industrial psychologists call the "shock absorber", that is, something which will introduce the Subject to the novelty of laboratory activities, but which will not be counted in his psychometric score. Wells also points out its weakness as a test for adults, although from the point of view of pictorial material it appeals to children. ('27, p. 126 f)

References:

Healy, '14 and '15
Kline and Kline '27
N. Y. Bureau of Analysis and Investigation '15
Pintner and Anderson, '17
Pintner and Paterson, '17
Gault '20
Skaggs, '20
Arthur, '25, and '28
Gaw, '25
Bronner, Healy, Lowe and Shimberg, '27
F. L. Wells, '27
Wallin, '27

No. 87 Pictorial Completion Test II

Devised by W. Healy
Date: 1921
Handled by Stoelting; also by Marietta.

Price: $15.00; (Stoelting); $15.00 (Marietta).

Material: Board and sixty small square pieces. The board shows eleven different scenes of experiences in a boy's day. In each picture there is a cut-out section.

Problem: To fit the most suitable pieces into the picture.

Individual test: Not adaptable to group testing.

Directions: Verbal; pantomime with difficulty.

Responses: Manipulative.

Time limit: Twenty minutes.

Scoring: Sum of values, assigned according to a table prepared by the author.

Standardization: This test has been standardized on 1542 cases by Healy and Bronner, for raw scores translated into age, and quartiles. Range: from seven years to adult.

Reliability: Bronner, Healy, *et al,* claim that this test is superior to pictorial completions of the Shaw type since the puzzle element has been eliminated. (27, p. 185)

Validity: r = .44 with Stanford-Binet on 282 fourteen-year olds. (Stolz) (Bronner, Healy, Lowe and Shimberg, '27, p. 186)
r = .33 with Stanford-Binet on a group of twelve-year olds.
r = .10 with Stanford-Binet on sixteen-year-olds. (F. L. Wells, '27, p. 129).

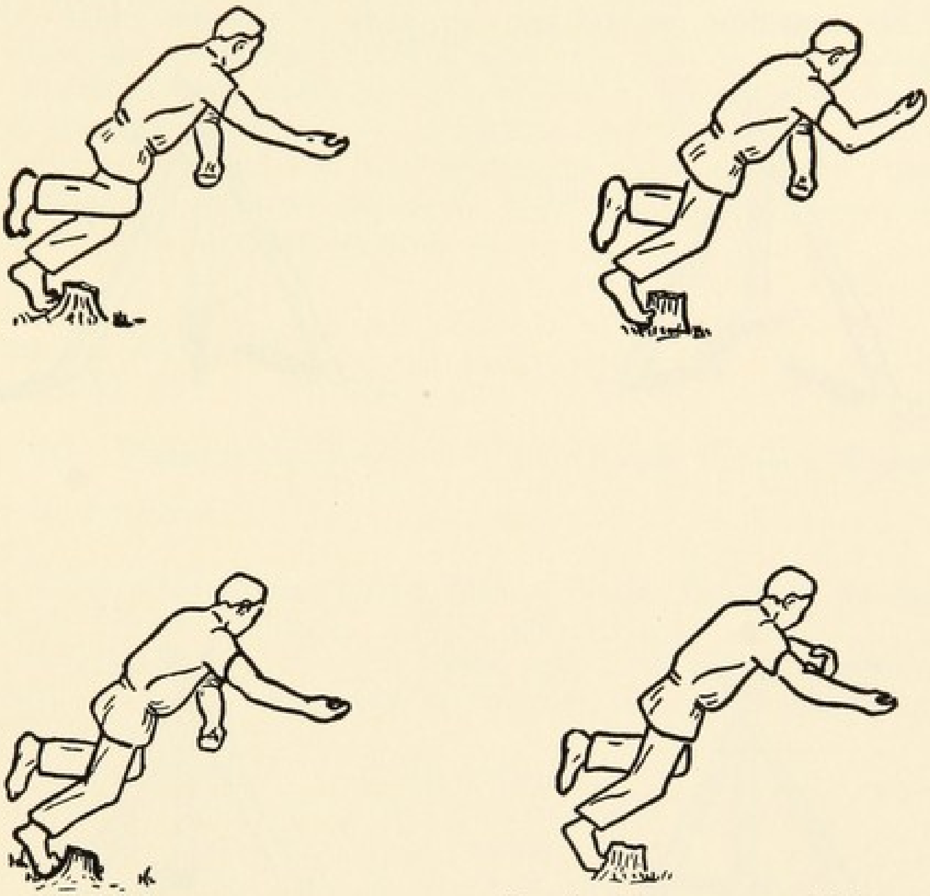
Discussion: Its authors point out that the main ability required for Pictorial Completion Test II is that of seeing relationships, of interpreting "the meaning of one element in the light of the others; to see relationships between the objects and the activities depicted and the missing pieces." (Bronner, Healy, Lowe and Shimberg '27, p. 185 f) Kent objects to this test because of its complicated scoring and because the pictures are not larger. She points out that the shadow item would be the most universally valid. (Personal interview.) F. L. Wells considers this test to be "among the major psychometric achievements". The continuity of situations, each with a missing element, calls for finer gradations of response in the selection of the cut-outs, as several possible choices involve fine discriminative judgments. Wells considers that this is a better test of "apperception" than is any other standardized test yet offered. He concludes that "common sense would rate this test high as a guage of general adaptive capacity, although systematic data are meager." ('27, p. 127 ff) Boody concluded from her use of a modified form of this test on Ellis Island immigrant children "that a test along this line well adapted for non-English speaking aliens .... could be evolved which, running smoothly and in silence, could show something of reasoning ability, without the intervention of pencil and paper". ('26, p. 132)

References:

Skaggs '20
Bronner '21 and '22
Healy '21
Morganthau '22
Perry '22
Gaw '25
Boody '26
F. L. Wells '27
Bronner, Healy, Lowe and Shimberg '27
Dorcus '28

No. 88 Picture Analysis Test

Devised by P. C. Squires.
Date: 1926
Produced in the Psychological Laboratory, Princeton University.



Reproduced by courtesy of Princeton Univ. Press

Materials: Exposure apparatus mounted on a base-board, 9 x 6 x 1 inches; fourteen groups of pictures, four pictures to a group. The pictures in each group are of the same object, but they differ from one another in certain details. (For details, see Squires '26, p. 89 ff)

Problem: Given a picture, to identify the duplicate of the picture among a group of four pictures, three of the four differing from the original in various qualitative respects; but all being of the same object.

Individual test: Series A, four practice items, twelve test items. Series B, twelve test items. This test is not adaptable to group testing.

Time: No time limit.

Scoring: Time, moves, successes.

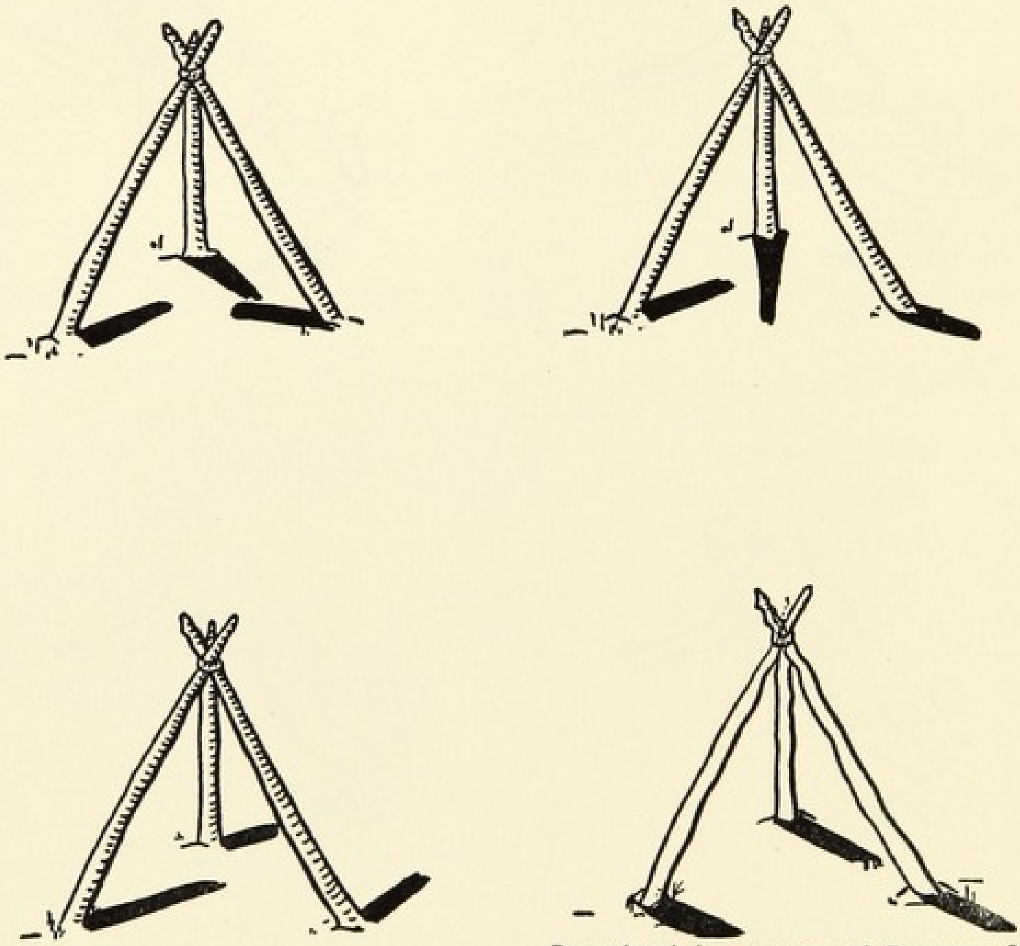
Directions: Pantomime.

Responses: Manipulative.

Standardization: Squires experimented with this test on fifty Princeton seniors and fifty feeble-minded males from Vineland. Norms are not yet established, but Squires considers that this test runs from lower levels to superior levels. (Squires '26, p. 106)

Validity: r = .02 (successes) and —.34 (time) for the Princeton group, criterion being the higher score on either the Thorndike or Princeton intelligence examination.

r = .31 (successes); and .19 (time) for the Vineland group, criterion being Stanford-Binet mental age. (p. 184)



Reproduced by courtesy of Princeton Univ. Press

Discussion: From the point of view of administration, this test "puts itself across". (p. 106) "The picture analysis test holds some good possibilities, especially in the province of racial psychology. The material utilized in this test is primitive and well-nigh universal, and appealed strongly to the type of mind found at Vineland." (Squires '26, p. 193)

References: Squires '26

No. 89 Picture Arrangement Test

Devised by A. C. Bowler and G. M. Whipple: modified by the Army psychologists.

Date: 1917

Published by no one at present.

Material: A series of five sets of "Foxy Grandpa" pictures placed out of order. one illustration set.
Problem: To place the pictures in their right order so as to make a sequence.
Individual test: five sets. This test is adaptable to group testing.
Directions: Verbal or pantomime.
Responses: Manipulative.
Time limit: Three minutes.
Scoring: Time and arrangement. One point for each pair in correct juxtaposition. Penalty of one point for each reversal. Time credited according to a table if arrangement is correct. (Yoakum and Yerkes '20, p. 115 f)
Standardization: This test was used and reported upon by the Army psychologists. Bowler ('17) had previously standardized it on "some 1012" cases, ranging in years from six to adult.
References:
Bowler, '17
Yoakum and Yerkes, '20
Yerkes, (Ed.) '21

No. 90 Picture Completion Test (Shaw Picture Puzzles)

Devised by E. A. Shaw.
Date: 1918
Handled by Psycho-Educational Clinic, Palfrey House, Harvard; also by Stoelting.
Price: $15.00 (Palfrey House); $15.50 (Stoelting).
Material: Wooden frame, 16 x 12 inches, with forty circular insets: ten correct responses, ten correct responses as to color, ten plain neutral colored, ten incorrect in every particular.
Individual test; not adaptable to group testing.
Problem: To fit the most appropriate inset into each depression.
Directions: Verbal: pantomime by adaptation.
Responses: Manipulative; made to situations and colors; insets are of identical size and shape, as are the depressions.
Time limit: Five minutes.
Scoring: Points according to a table. Score ranges from minus thirty to plus one hundred for a perfect performance. (Dearborn *et al*, '23, p. 35)
Standardization: Shaw reports raw scores based on ninety cases, distributed by ages. Range; six to twelve years. (Dearborn *et al*, '23, p. 609)
Reliability: No data as yet.
Validity: No data as yet.
Discussion: Bronner, Healy, *et al* contend that this test is characterized by a puzzle element which makes it undesirable for intelligence testing ('27, p. 185). But Shaw, (in a letter to the writers) maintains that the test involves a problem in selection rather than a puzzle. "The test was constructed to offer a picture characterized by a unity not found in tests of a like nature. Secondly, the effort was made to produce a test which would be superior in its mechanical aspects to any which had at that time appeared." (Dearborn, Shaw and Lincoln, '23, p. 33 f) F. L. Wells states that some people prefer the square cut-outs of the Healy picture completion tests to

the circular ones in this. But since the Shaw test is drawn on a larger scale, it is clearer than either of the Healy tests, and, from the standpoint of "complexity of situations it stands between" them. ('27, p. 127)

References:

Shaw '18, and letter to writers
H. L. Hollingworth, '20
Dearborn, Shaw and Lincoln, '23
F. L. Wells, '27

No. 91 Ship Test

Devised by B. Glueck; modified by H. A. Knox.
Date: 1914
Handled by Stoelting; also by Marietta.
Price: $5.00 (Stoelting) $4.50 (Marietta).
Material: Picture of a ship mounted on a board and cut into ten equal-sized pieces, all similar in shape. Pintner and Paterson used a frame, 25 x 16.2 centimeters on the outside, and 21.4 x 12.5 centimeters on the inside. Into this they fitted the pieces.
Problem: To fit the pieces together in their correct arrangement.
Individual test; adaptable to testing in small groups.
Directions: Verbal or pantomime.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Points in terms of placings.
Standardization: This test was standardized by Pintner and Paterson on 648 school children. Norms are expressed as raw scores, distributed for age, percentiles and quartiles. Range; five to fourteen years. This test discriminates well from five to eleven years. (Pintner and Paterson '17).
Reliability: "The variability at most ages is not very great and the quartile diminishes markedly in the upper ages." (Pintner and Paterson '17, p. 132)
Discussion: "The subject must be guided solely by the picture he is trying to make." (Pintner and Paterson '17)
References:

Knox '14
Pintner and Paterson '17
Yoakum and Yerkes '20
Baldwin and Stecher '25
F. N. Freeman '26
Bronner, Healy, Lowe and Shimberg '27

No. 92 Special Picture Puzzle

Devised by W. Healy and G. M. Fernald.
Date: 1911
Handled by Stoelting.
Price: $5.50
Material: Colored picture of a school room, showing eight pupils and a teacher. Thirteen cut-outs representing definite unit parts in the scene.

Problem: To put the pieces in the board as quickly as possible.
Individual test only; not adaptable to group testing.
Directions: Verbal.
Responses: Manipulative.
Time: Average about three minutes.
Scoring: Time; errors; absurd errors.
Standardization: This test was never adequately standardized. Healy early reported that kindergarten children were able to do it. It was too simple for older children and hence discarded in favor of later completion tests. (Healy '15, p. 107)
Discussion: This test is included here merely in the interest of comprehensiveness and because it is one of the earliest pictures completion boards on record. It is doubtful if it will ever be given preference over later models.
References:
Healy and Fernald, '11
Schmitt, '15

No. 93 Geographical or Jig Saw Test

Devised by H. A. Knox.
Date: 1914
Handled by Stoelting.
Price: $4.50
Material: Rectangular board on which is painted a geographical scene. Six irregular sections are cut out of the margin.
Problem: To place the pieces into their proper clefts.
Individual test; not adaptable to group testing.
Directions: Verbal or pantomime.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time and improvement time.
Standardization: This test was first reported by Knox, who places it at seven years. It was suggested for immigrants above nine years by Mullan, who gives a distribution of scores for time and errors for literate and illiterate men and women. (Mullan p. 85f)
Discussion: The test was never adequately standardized and has apparently fallen into disuse. Mullan noted that while the literate groups in general did better than the illiterate, nevertheless, the results were not always consistent. He found it interesting to know their procedure and general method of attack, to detect evidences of reasoning or planning, or just trial and error performance.
References:
Knox '14
Mullan '17

VIII

CONSTRUCTION TESTS

No. 94 Construction Test A

Devised by W. Healy; modified by G. Fernald and W. F. Dearborn.

Date: 1911

Handled by Stoelting.

Price: $2.50

Material: Wooden frame, 15.3 x 12.7 centimeters; and five rectangular pieces.

Problem: To see how quickly the pieces can be fitted into the frame.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time limit: Five minutes.

Scoring: Number of moves, (correct and incorrect noted); time in seconds.

Standardization: This test was standardized by Lowe, Shimberg and Wood, on 1596 cases. Norms are expressed for time and moves, in terms of age, percentiles, and quartiles, ranging from nine to seventeen years. (Pintner and Paterson, '17)

Reliability: r = .23 for time, first and second trials. r = .21 for time, A against B, (Morganthau, '22) r = .15 for time, A against B. (Bronner, Healy, Lowe, Shimberg, '27)

Validity: r = .19 against Stanford-Binet I. Q. "But from the data above we see no evidence for a reliable indication of either general intelligence or any special ability." (Bronner, Healy, Lowe and Shimberg, '27, p. 217)

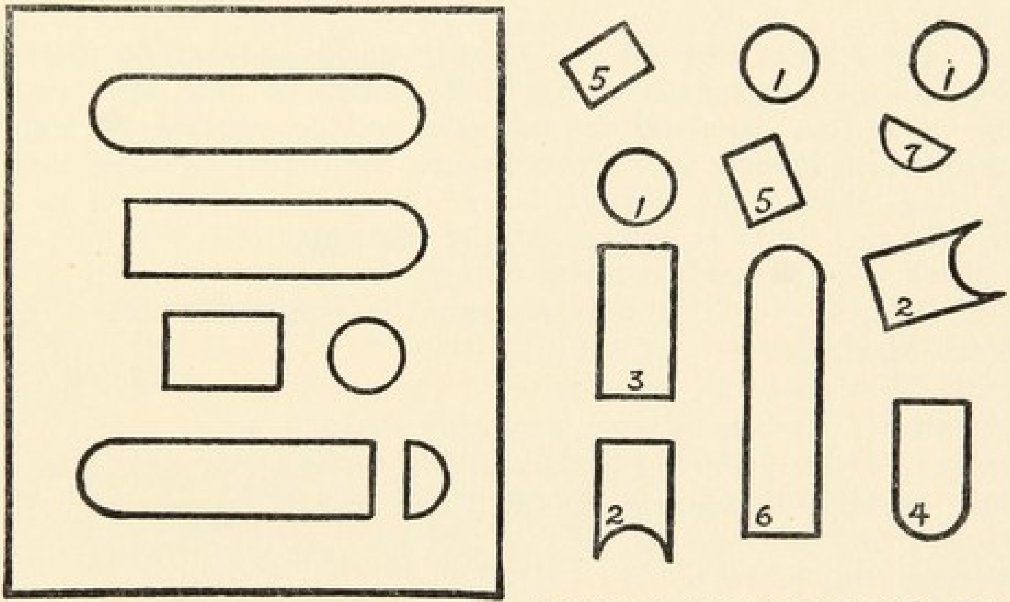
Discussion: Bronner, Healy, *et al,* suggest that this test has value in observing the individual's approach, method of attack, and the like. They suggest limiting its use to a general battery. According to F. L. Wells, Bronner considers this test as one of ability to solve particular kinds of problems, involving relationships regardless of age. Wells thinks that Healy A is the best of the construction puzzles. Later, in discussing the discrepancy in norms between the Pintner, Paterson and the Lowe, Shimberg, Wood, standardization, Wells suggests that this discrepancy could be accounted for, but inadequately, by differences in group selection or administrative variations. He adds that "this difference throws into strong relief the coarseness of these tests and the impossibility of drawing refined distinctions on the basis of single performances." ('27, p. 131) Drever and Collins also have produced a modification of this test.

References:

Healy and Fernald, '11
N. Y. Bureau of Analysis and Investigation '15
Schmitt, '15
Healy, '15
Terman, '16
Weidensall, '16
Bruckner and King, '16
Pintner and Paterson, '17
Morganthau, '22
Lowe, Shimberg and Wood, '24
Gaw, '25
F. L. Wells, '27
Bronner, Healy, Lowe and Shimberg '27
Drever and Collins, '28

No. 95 Construction Test B

Devised by W. Healy; modified by G. Fernald and W. F. Dearborn.
Date: 1911
Handled by Stoelting.
Price: $6.50.



Reproduced by courtesy of Little, Brown & Co.

Material: Frame of wood, and eleven pieces.
Problem: To see how quickly the pieces can be put into their proper places.
Individual test; not adaptable to group testing.
Directions: Verbal or pantomime.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time in seconds and numbers of moves. Record placings by number.
Standardization: This test was standardized by Lowe, Shimberg and Wood on 1596 cases. Norms are expressed for time and moves in terms of age and quartiles. Range: Nine to seventeen years.
Validity: r = .27 against Standford-Binet I. Q. (Bronner, Healy, Lowe Shimberg, '27, p. 217) This low correlation is substantiated by Morganthau's data.
Discussion: This test is best limited for use in a general battery. Like "Healy

A", it affords an opportunity to observe an individual's method of attack and procedure. Its authors do not consider it a good test of intelligence, since its solution is dependent on the element of chance. (Bronner, Healy, Lowe and Shimberg, '27, p. 216)

References:

Healy and Fernald, '11
Healy, '15
Schmitt, '15
New York Bureau of Analysis and Investigation '15
Weidensall, '16
Morganthau, '22
Lowe, Shimberg and Wood, '24
Bronner, Healy, Lowe and Shimberg '27

No. 96 Construction Puzzles

Devised by H. T. Woolley.

Date: 1926

Handled by Stoelting.

Price: $7.50.

Material: The basis of this test is a game or puzzle supplied by toy stores, and known as the "egg of Columbus." It consists of nine little stone blocks, made up in four pairs and one odd one, and six patterns cut out of three-sixteenth inch binding board, to represent the following shapes: egg, flower-pot, chick, ship, cradle and seal.

Problem: To put the stone blocks into the patterns.

Individual test; not adapted to group testing.

Directions: Verbal; might possibly be given by pantomime.

Responses: Manipulative.

Time limit: Five minutes.

Scoring method: Median between percentile ranks.

Standardization: This test was standardized by Woolley on 1432 male and 1212 female school and working children, ages sixteen to eighteen. Norms are expressed for time in terms of age and sex and distributed by percentiles. (Woolley, '26, p. 140)

Discussion: "The tests are good ones of their type, but it is hard to say what their significance is. Whatever they measure, it seems unrelated to school grade, or to the contrast between working and school groups, to a greater extent than any other mental test." (Woolley, '26, p. 141) Bronner, Healy, Lowe and Shimberg suggest that an inherent defect in the test or the use of an inadequate time limit may account for the many failures, even at the eighteen-year level. ('27, p. 217)

References:

Woolley, '26
Bronner, Healy, Lowe and Shimberg, '27

No. 97 Diamond Frame Test. (Diagonal Frame Test)

Devised by W. Healy, G. Fernald and F. Freeman as "Construction A;" modified by H. A. Knox.

Date: 1913

Handled by Stoelting.
Price: $3.25
Material: A wooden frame cut "on the bias," similar to "Healy A"; with six sections.
Problem: To fit the pieces into the frame as quickly as possible.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Five or ten minutes.
Scoring: Time and improvement time. (Knox, '14)
Standardization: This test was reported by Knox, who, however, did not standardize it. Knox scaled it at nine years.
Reliability: The Diamond Frame, with the Knox Moron, together correlate against Healy A and B: r = .161 for time; r = .76 for moves. (Morganthau, '22, p. 45)
Discussion: This test belongs to the type that has fallen into discard. It has drawn little or no attention to itself since Knox offered it, and Morganthau's study has confirmed its inadequacy.
References:
Knox, '13 and '14
Morganthau, '22

No. 98 Diagonal Test

Devised by G. A. Kempf.
Date: 1914
Handled by Stoelting.
Price: $2.75
Material: Wooden frame, 16.5 x 12.7 x 1 centimeters; and five pieces.
Problem: To fit the pieces into the frame as quickly as possible.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring: Time and errors.
Standardization: This test was standardized by Pintner and Paterson on 619 school children. Norms are expressed for time and errors in terms of age, percentiles and quartiles. Range: from five to thirteen years.
Reliability: Pintner and Paterson point out that the test offers different ways of fitting in the pieces, all of which ways are not equally difficult. The child who gets started on one of the easier ways has the advantage over another who has made a harder but equally sensible initial move. This element of chance places the Diagonal Test in the "puzzle type of test". ('17, p. 42f)
Discussion: Pintner and Paterson consider this test worth while only when used in a group of tests. ('17, pp. 119-121) F. L. Wells thinks that on account of its resemblance to the "Healy A", it may be interchanged with the latter, although he points out that the Diagonal Test appears to be slightly easier than the "Healy A." ('27, p. 136)

References:

Knox, '14
Pintner and Paterson, '17
Johnson, '25
Gaw, '25
F. L. Wells, '27
Bronner, Healy, Lowe and Shimberg '27

No. 99 Feature Profile Test

Devised by H. A. Knox; modified by G. A. Kempf; again modified by Pintner and Paterson.

Date: 1914

Handled by Stoelting; also by Marietta.

Price: Knox-Kempf modification $5.00; Pintner-Paterson modification $3.75 (Stoelting); $3.50 (Marietta).

Material: Wooden board, 21 x 17 x 1 centimeters, shaped like the profile outline of a human head, and seven small pieces. (Pintner and Paterson, 17)

Problem: To construct profile of the head from the parts given.

Individual test; not adaptable to group testing.

Directions: Verbal or pantomime.

Responses: Manipulative.

Time limit: Five minutes.

Scoring: Time. Any errors constitute total failure.

Standardization: This test was revised and standardized by Pintner and Paterson ('17) on 713 school children, ranging from six to sixteen years. Norms are reported for time and expressed in terms of age, percentiles, and quartiles. The test does not discriminate above eleven years.

Reliability: "The variation of the middle fifty percent is rather large, and varies considerably in amount for the ages tested." (Pintner and Paterson, '17, p. 129) The test is considered unsatisfactory "because the individual's performance is thrown out if the four pieces forming the ear are fitted in without regard for their proper relation to each other." (Bronner, Healy, Lowe and Shimberg, '27, p. 219)

Discussion: "It demands the synthetic ability of seeing the parts of a whole and of putting these together, a kind of ability which seems to be one of the essential factors in general intelligence." (Pintner and Paterson, '17, p. 56)

References:

Knox, '14
Pintner and Paterson, '15 and '17
Yoakum and Yerkes, '20
Yerkes, (Ed.), '21
Arthur, '25, '28
Bronner, Healy, Lowe and Shimberg '27

No. 100 Graded Series of Geometrical Puzzles

Devised by G. H. Kent.

Date: 1916

Material: Twenty-four puzzles consisting of rectangular or triangular pieces of wood. Each puzzle, when solved, forms a square.

Problem: To fit the pieces together into the frame. Examiner alternates between rectangles and triangles to reduce the practice effect.
Individual test.
Discussion: This series has been definitely discarded by its author, who wishes it to be considered obsolete.
References:
Kent, '16 b, and letter to writers.

No. 101 Imbecile Test

Devised by H. A. Knox
Date: 1914
Handled by Stoelting.
Price: $4.50
Material: Wooden frame, 5 x 7 x 1 inches, with nine cut-outs, numbered.
Problem: To fit the pieces into the board as quickly as possible.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Credit, if less than six mistakes.
Standardization: This test was reported by Knox who offered it for age six years, but who made no claim to have standardized it.
Discussion: This test was tried out by Pintner and Paterson, but rejected by them because it did not seem to be giving satisfactory norms. ('17, p. 23)
References:
Knox. '14
Pintner and Paterson, '17

No. 102 Knox Moron Test

Devised by W. Healy; modified by H. A. Knox.
Date: 1914
Handled by Stoelting.
Price: $2.50
Material: Wooden frame, 4½ x 4 x 3/5 inches, with four cut-outs; numbered.
Problem: To fit the blocks into the frame.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Individual test, three trials. This test is not adaptable to group testing.
Time: Knox allowed ten minutes, which is apparently too long.
Scoring: Time and improvement-time over earlier trials. (Knox, '14)
Standardization: Knox set this test at ten years, but did not claim to have standardized it. It has drawn very little attention, and Morganthau's results have supported its neglect.
Reliability: $r = .135$ with "Healy A" for time.
$r = .27$ with Knox Diamond Frame, for time.
$r = .103$ with Diamond Frame, for moves.

r = .161, average of Healy A and B with Knox A and B, for time.
r = .076, average of Healy A and B with Knox A and B, for moves. (Morganthau, '22, p. 45)

Validity: r = .046 ± .095 for time, on 111 cases with Stanford-Binet.
r = .009 ± .099 for moves, on 103 case with Stanford-Binet. (Morganthau, '22, p. 38).

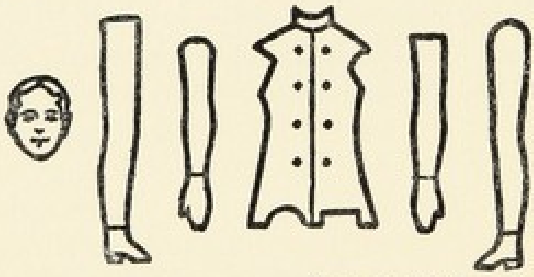
Discussion: This test, and those which precede, of the construction puzzle type, were offered originally as tests of general ability, but as F. L. Wells points out, they were found unsatisfactory in clinical practice as age-level criteria. The difficulty of wide variation in scores is always encountered, making a single observation based on tests of this type of but little diagnostic value. ('27, p. 131)

References:

Knox, '14	F. L. Wells, '27
Morganthau, '22	Davenport, '29

No. 103 Manikin Test

Devised by R. Pintner.
Date: 1917
Handled by Stoelting; also by Marietta.
Price: $3.60 (Stoelting); $4.00 (Marietta).



Reproduced by courtesy of Little, Brown & Co.

Materials: Six small wooden pieces representing the conventional figure of a man: body, two arms, two legs, head.

Problem: To construct the figure of a man from the parts given.

Individual test; not adaptable in its concrete form to group testing.

Directions: Verbal or pantomime.

Responses: Manipulative.

Time limit: Five minutes.

Scoring: Points according to a table of directions. (Pintner, Paterson, '17, p. 55) Norms are distributed for age, in percentiles and quartiles.

Standardization: This test was standardized by Pintner and Paterson, ('17) on 545 school children, ranging from four to ten years. They found that the test did not discriminate above eight years, but was excellent for differentiating abilities below age eight.

Discussion: "It demands the same kind of ability as the Feature Profile test. The scattered fragments suggest some kind of a complete whole and the child has to synthesize these scattered impressions and plan to reach a definite end." (Pintner and Paterson, '17, p. 53) Stutsman found this test unsatisfactory. ('26)

References:

Pintner and Paterson, '17
Yerkes, '20
Yoakum and Yerkes, '20
Johnson, '25
Baldwin and Stecher, '25
Arthur, '25 and '28
Stutsman, '26
Bronner, Healy, Lowe and Shimberg '27
Davenport, '29

No. 104 Triangle Test

Devised by M. K. Gwyn.
Date: 1914
Handled by Stoelting.
Price: $2.75
Material: Wooden frame, 17 x 12.8 x 1 centimeters; four right-angled triangles, 6 x 9.5 centimeters.
Problem: To fit the pieces into the frame as quickly as possible.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time and errors.
Standardization: This test was standardized by Pintner and Paterson on 620 school children. Norms are expressed for time and errors in terms of age, percentiles and quartiles. Range; five to thirteen years. The test is considered especially suitable for children below eight years.
Reliability: The irregular score scattering seems "to be due to the puzzle nature of the test which allows for the entrance of a chance solution every now and then." (Pintner and Paterson, '17, p. 118) F. L. Wells also mentions that this test offers large possibilities of chance success. ('27, p. 136)
Discussion: Bronner, Healy, *et al*, say that the test is useful only for children below eight years of age. ('27, p. 222)

References:

Gwin, '14
Knox, '14
Pintner and Paterson, '17
Gaw, '25
Bronner, Healy, Lowe and Shimberg '27
F. L. Wells, '27

IX

FORM BOARDS

No. 105 Adaptation Board

Devised by H. H. Goddard; modified by Pintner and Paterson.
Date: 1912
Handled by Stoelting.
Price: $4.50.
Material: Board with four circular holes, one larger than the other three; a circular wooden disc to fit the largest hole. Goddard board, 22 x 28 centimeters, with holes 6.3 and 6.5 centimeters in diameter. Pintner-Paterson board, 22 x 25 x .5 centimeters, with holes 6.8 and 7 centimeters in diameter. Wooden block with handle to fit the large hole exactly.
Problem: To fit the disc into the right hole.
Individual test; five positions on the board. This test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: A minute or two.
Scoring: Number of moves.
Standardization: This test was standardized by Pintner and Paterson, ('17) on 827 school children. Norms are expressed for moves in terms of age, percentiles and quartiles. Range: five to thirteen years; discriminative up to ten years of age.
Discussion: Bronner, Healy, *et al*, feel that this test is useful for children below age eight. They state that Goddard himself believes that it has distinct value only when included in a scale of tests. ('17, p. 222)
References:
Goddard, '12a and '15
Pintner and Paterson, '17
Gaw, '25
Bronner, Healy, Lowe and Shimberg, '27

No. 106 Arrow Board

Devised by F. L. Dunham.
Date: 1916
Handled by Stoelting.
Price: $12.75
Material: Wooden board with arrow-shaped cut-out; and ten pieces.
Problem: To fit the pieces into the board as quickly as possible.
Individual test; three trials. This test is not adaptable to group testing.
Directions: Verbal; or pantomime by adaptation.

Responses: Manipulative.
Time: Less than one minute.
Scoring: First attempt; describe quality of response. Second and third attempts, record time in seconds.
Standardization: Dunham reported his work with this test on 184 high school juniors and seniors. Norms are expressed in terms of time, for ages fifteen to eighteen. Dunham offers this test for "children, low grade adults and illiterate aliens," but Pintner and Paterson say that "the number of subjects tested and the nature and selection of subjects, make the test of little value in practical clinical work at the present time." ('17, p. 95)
Discussion: Dunham finds the chief value of this test in the "subject's ability to differentiate circles, rectangles and triangles, as steps in development of visual analysis". Insight, co-ordination, poise and use of one or both hands are also suggested in Dunham's article. ('16, p. 286)
References:
Dunham, '16
Pintner and Paterson, '17, (p. 95)
Bronner, Healy, Lowe, Shimberg, '27

No. 107 Casuist Form Board

Devised by H. A. Knox; modified by Sprague (?)
Date: 1914; revised 1917
Handled by Stoelting; also by Marietta
Material: Price $6.50 (Stoelting); $7.00 (Marietta)
Material: The original Sprague-Knox board is 13 x 10 x ½ inches. The Pintner-Paterson board is 50 x 25.7 x 1.5 centimeters. Both have three circular recesses and one elongated oval cut-out; with blocks, of which the two larger circles are each cut into three equal segments, and the smaller circle is cut into two equal segments. The oval is cut into four pieces, two circular end pieces and two middle pieces, making in all, eleven pieces.
Problem: To fit the pieces into the board as quickly as possible.
Individual test. This test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time and number of errors.
Standardization: Norms are expressed for time, moves and errors in terms of age, percentiles, and quartiles. The test has been standardized on 918 school children by Pintner and Paterson. Range: five through fourteen years, but not discriminative above ten or eleven years of age.
Discussion: Pintner and Paterson considered the distribution on this test to be good, on the whole (p. 112) but Arthur found the test to be one of the least satisfactory in her scale. ('25)
References:
Knox, '14
Pintner and Paterson, '17
Johnson, '25
Arthur, 25, '28
Bronner, Healy, Lowe, Shimberg, '27
Riley, '29

No. 108 Cylinder Test

Devised by L. Witmer. Based on Montesorri materials.
Date: 1916
Handled by the Psychological Laboratory and Clinic, University of Pennsylvania; also by Stoelting.
Price: $35.00 (Stoelting).
Material: A circular board, "having a series of recesses about its outer edge into which are fitted eighteen cylinders of different depths and diameters, with a central compartment, into which the blocks can be thrown."
Problem: To put the blocks into their proper holes.
Individual test: Three trials; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes for each trial.
Scoring: Final score is time on shortest trial.
Standardization: This test was standardized by Paschal on 2230 school children. Norms are expressed for time in terms of age, sex, percentiles and quartiles. Range: from six years to adult.
Validity: Paschal found an increase in Cylinder Test performance with increasing pedagogical age, a direct relation between such performance and general proficiency in daily activities, and a rather high positive correlation with manuai or mechanical capacity, as measured by shop ratings. (Paschal, '18b, p. 47-51).
Discussion: Several investigators have stressed the value of this test for analytical diagnosis. (A. M. Jones '25, Leaming '23, Paschal '18) The Subject's method of attacking the problem, as well as his "planfulness, his image-ability, his powers of analysis and his powers of distributive attention" are all indicated to some extent by his performance on the Cylinder Test. (Bronner, Healy, Lowe and Shimberg, '27, p. 218)
References:
H. H. Young, '16
Ide, '18
Paschal, '18a and '18b
Leaming, '23
Woolley and Cleveland, '23
Easby-Grave, '24
Paschal, '25
A. M. Jones, '25
Bronner, Healy, Lowe and Shimberg, '27

No. 109 Ferguson Form Boards

Devised by G. O. Ferguson, Jr.
Date: 1921
Handled by Stoelting.
Price: $58.50 for complete set of six boards.
Material: Six graded form boards, made of wood, each containing six spaces, some to be filled by one, some by two blocks; some with beveled edges, some with grooved edges. Size, when stacked together, about one cubic foot.
Problem: To fill in spaces with blocks according to directions. Six problems.
Individual test; not adaptable to group testing.

Directions: Verbal, or pantomime, but usually the former.
Responses: Manipulative.
Time: About fifteen minutes, but varies greatly with the Subject. Limit of five minutes for each board.
Scoring: Points according to a table of time values.
Standardization: This test was standardized by several investigators on various groups of from two to three thousand cases. (See Bronner, Healy, *et al* ('27, p. 127): 308 males, 171 females, (Meehan, Shimberg) Norms are expressed for raw scores and time in terms of age, grade and sex; distributed in percentiles and quartiles. Range: four years to no limit. This test is especially applicable to pre-school children, primitive peoples, and superior adults. (Ferguson, letter to writers).
Reliability: $r = .80$ (Ferguson); $.90 \pm .02$ retest. (MacPhee and Brown '30, p. 29)
Validity: $r = .70$. Criterion: school grades. (Ferguson letter to writers).
$r = .33 \pm .06$ Criterion: N. I. T.; 75 cases. (MacPhee and Brown '30)
$r = .13 \pm .09$ Criterion: Pintner-Cunningham Primary Group Test, 27 cases. (MacPhee and Brown '30, p. 29).

MacPhee and Brown ('30), who tried this test out on 134 school children, aged eight to twelve years, report a wide variability at each age. They state that their scores showed no significant increments at successive age levels, and even by trying an alternative method of scoring, found very large deviations for the increments that did appear.

Discussion: This test, according to its constructor, tests "formboard ability." He thinks it could be used on peoples of different cultures and background. "Much material from the use of this test has been accumulated, but has not yet been analyzed. Improved scoring systems and norms will be devised if possible." (Ferguson, letter to writers). MacPhee and Brown consider that the test has value in offering for analysis certain qualitative data on behavior during examination which can be scored objectively. The data on such points as method of attack, procedure, attitude, and so forth, show significant differences for high and low score cases, which probably accounts in part for differences in scores. ('30, p. 35)

References:
Schmitt, '15
Ferguson, '20
McFarlane, '24
Brotemarkle, '27, '28
Bronner, Healy, Lowe and Shimberg, '27
MacPhee and Brown, '30

No. 110 Five Figure Board

Devised by D. G. Paterson.
Date: 1917
Handled by Stoelting; also by Marietta.
Price: \$6.50 (Stoelting); \$7.00 (Marietta).
Material: Formboard, 57.4 x 20.3 x 1.2 centimeters. Formboard has five cut-outs, each piece being cut into two irregularly shaped pieces, and one being cut into three.

Problem: To fit the pieces into the board.

Individual test; not adaptable to group testing.

Directions: Verbal or pantomime.

Responses: Manipulative.

Time limit: Five minutes.

Scoring: Time and errors.

Standardization: This test was standardized by Pintner and Paterson on 933 school children. ('17) Norms are expressed in terms of time and errors; and distributed for ages, percentiles and quartiles. Range: from five to fourteen years.

Discussion: Pintner and Paterson found this test distinctly harder than the Seguin formboard. They also found an element of the puzzle-board idea in it, which may influence the performance of the test to a slight extent. (Pintner and Paterson, '17, p. 105f)

References:
Pintner and Paterson, '17
Bronner, Healy, Lowe, Shimberg, '27

No. 111 Form Board 1-A

Devised by W. F. Dearborn and J. E. Anderson.

Date: 1916

Handled by Psycho-Educational Clinic, Harvard University.

Price: $6.00. Manual; with other information: Seventy-five cents.

Material: Formboard, 17.5 x 11 x 1 inches, into which eight depressions have been cut; sixteen pieces. Each block fills half a depression, divided lengthwise.

Problem: To fill up empty spaces, each with two blocks, as quickly as possible.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Limit of five minutes.

Scoring: Number and kind of moves; and time.

Standardization: Dearborn, Shaw and Lincoln have offered a tentative standardization, based on 433 school children. Norms are expressed for time and moves in terms of age. Range: from five to ten years. ('23, p. 41)

Discussion: Dearborn states that this formboard tests "form concepts and manipulative skill". (letter) F. L. Wells considers that the Dearborn formboard involves appreciation of complex relationships, which calls for more planning than does the "construction puzzle." ('27, p. 130)

References:
Dearborn, Anderson, Christianson, '16
Dearborn, Shaw and Lincoln, '23
Bronner, Healy, Lowe, Shimberg, '27
F. L. Wells, '27

No. 112 Form Board 1-C

Devised by W. F. Dearborn and J. E. Anderson.
Date: 1916
Handled by Psycho-Educational Clinic, Palfrey House, Harvard. Stoelting handles a board which is a combination of Dearborn 1-A and 1-C, having a single base board for both.
Price: $6.00 (Palfrey House); $16.50 (Stoelting).
Material: Formboard 17.5 x 11 x 1 inches, into which eight depressions have been cut; twelve blocks as in Form Board 1-A with each of the four remaining blocks cut in half.
Problem: To fill up spaces by moving around the blocks in the partially filled spaces, with as few moves as possible.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Scoring: Time and moves.
Standardization: Dearborn, Shaw and Lincoln offer a tentative standardization of this test, based on examination of 433 school children. The scores, time and moves, are distributed according to age. Range: from five to ten years. ('23, p. 43)
Discussion: Like Formboard 1-A, according to its makers, this is a test of "form concepts and manipulative skill." Dearborn *et al* doubt the validity of a standardization based on number of correct and incorrect moves and time needed to complete the task. ('23, p. 5) Leaming also points out a possible fallacy, or at least, unfairness in this method of scoring, in that it penalizes the slower but more painstaking individual who takes time to plan his move; while the rapid individual who, by an inferior method, makes several trial and error moves, finally stumbles upon the correct one and gets through in less time. "Pertinent differences which the trained observer recognizes are not gauged merely by the time taken or the number of moves made by the subject". (Dearborn *et al*, '23, p. 19)
References:

Dearborn, Anderson Christianson, '16
Leaming, '22
Dearborn, Shaw and Lincoln, '23
Boody, '26
Bronner, Healy, Lowe, Shimberg, '27
Murphy, '28
Brooke, '28
Bidelle, '28
Farson, '28

No. 113 Form Board 2. (Reconstruction Puzzle)

Devised by W. F. Dearborn and J. E. Anderson.
Date: 1916
Handled by Psycho-Educational Clinic, Palfrey House, Harvard; also by Stoelting.
Price: $4.50 (Palfrey House); $6.50 (Stoelting) Manual: Seventy-five cents.
Material: Form board 12 x 12 x ½ inches with eight irregular depressions; three blocks.
Problem: To fit the three blocks into each of the depressions.
Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Twenty minutes or less.
Scoring: Time.
Standardization: This test is not standardized, although results on thirty cases are offered. Range: four to sixteen years. (Dearborn, Shaw, Lincoln, '23, p. 46)
Discussion: A modification, 2-B, having four instead of eight depressions, has been adapted for younger children. (Dearborn *et al*, '23, p. 25)
References:
Dearborn, Anderson, Christianson, '16 Dearborn, Shaw, Lincoln, '23

No. 114 Form Board 3. (Block Test)

Devised by W. F. Dearborn.
Date: 1916
Handled by Stoelting.
Price: Wood, $10.00. Aluminum, $15.00. Manual, Seventy-five cents.
Material: Form board, 13 x 11½ x ½ inches, or 10 x 11½ x 1 inches; and twenty-one blocks, arranged four different ways according to pattern set by authors.
Problem: After demonstration, Subject is required to fill spaces which are already partly filled, as quickly as possible, and with as few moves as possible.
Individual test; three different problems. This test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Fifteen minutes or less.
Scoring: Time and moves.
Standardization: Dearborn, Shaw and Lincoln offer tentative standards based on 362 school children. Norms are distributed for time and moves, according to age. Range: six to thirteen years. ('23, p. 47f)
Discussion: The authors consider this a test of "form concepts and manipulative skill." It was employed as Test 5 of the "Performance Scale," in the Army tests.
References:
Dearborn, Anderson and Christianson, '16
Yoakum and Yerkes, '20, p. 108 f
Yerkes, (Ed.), '21, p. 185
Dearborn, Shaw, Lincoln, '23
F. L. Wells, '27
Bronner, Healy, Lowe and Shimberg, '27

No. 115 Form Board 4. (Triangle Performance Test)

Devised by W. F. Dearborn. This is a modification of the Triangle Test reported by Dearborn, Anderson and Christianson ('16).
Date: 1916
Handled by Psycho-Educational Clinic, Palfrey House, Harvard University; also by Stoelting.
Price: $6.00 (Palfrey House); $15.00 (Stoelting).

Material: Form board, $11\frac{1}{2} \times 14\frac{1}{2}$ inches, with twelve recesses. Two triangular insets.

Problem: To fit the two small triangles successively into each of the twelve depressions in the board.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Limit of ten minutes.

Scoring: Time.

Standardization: This test is not yet adaquately standardized, but has been reported upon by Dearborn, Shaw and Lincoln for 106 children, ages seven and eight years. The test is suitable for ages five years up, the upper limit not yet determined. ('23, p. 55)

Discussion: This test was first offered as "Triangle Performance Test" in 1916, when its authors advocated it primarily as an alternate test. (Dearborn, Anderson and Christianson '16, p. 455). It tests largely "form percepts and manipulative skill." Drever and Collins also have modified this test for use in England.

References:

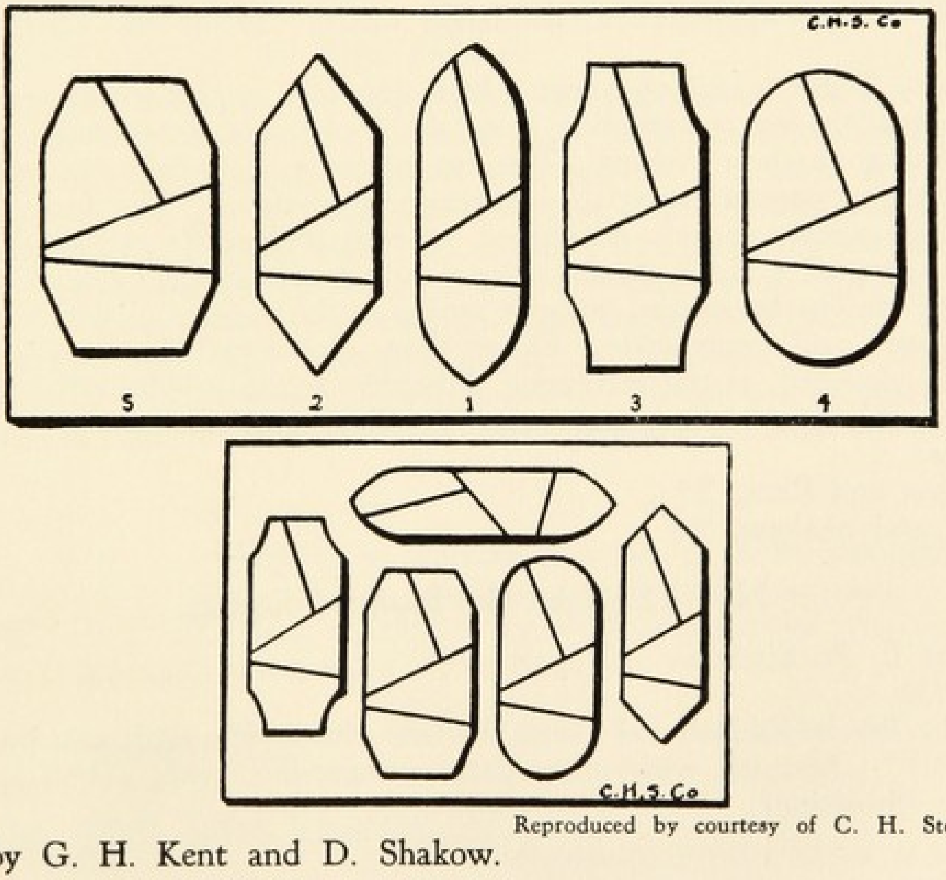
Dearborn, Anderson and Christianson, '16

Dearborn, Shaw, Lincoln, '23

Bronner, Healy, Lowe and Shimberg, '27

Drever and Collins, '28

No. 116 Graded Series of Form Boards



Reproduced by courtesy of C. H. Stoelting Co.

Devised by G. H. Kent and D. Shakow.

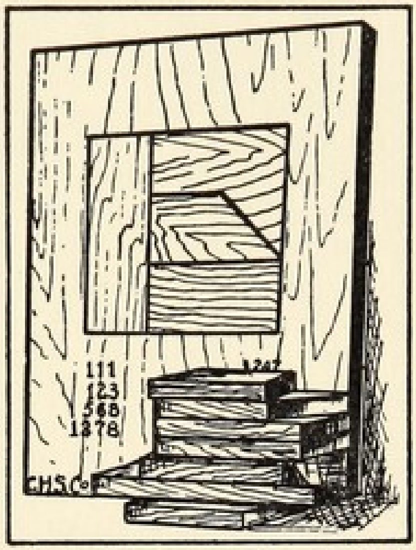
Date: 1925; date of revision 1928
Handled by Stoelting; also by Ellen Wilson, Rome, N. Y.
Price: $60.00 for industrial model; $54.00 for clinical model. (Stoelting).
Materials: There are two models which differ in size but not in principle. The industrial model is 22 x 10 inches, with five recesses, each of which is fitted with seven complete sets of blocks a little thicker than the depth of the recess. The clinical model, frame, twelve inches in length, is the same as industrial model, but reduced in size and differing slightly from the larger model in the proportion of the depressions as well as in their arrangement. This material is a revision of The Worcester Form Board series.
Problem: The seven sets of blocks furnish eight tasks, in each of which the contents of all five recesses are divided in exactly the same way. Each task involves one specific problem which is presented five times.
Individual test; eight tasks graded in difficulty. This test could not be adapted to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Twenty to thirty minutes.
Scoring: Time records are converted into point values, but no longer according to table published by Kent and Shakow. ('28, p. 119) Revised scoring method furnished by Dr. Kent on application.*
Standardization: The industrial model has not been standardized. The clinical model has been tentatively standardized by Kent on 100 children. Range: five years to superior adult. Norms furnished on application.* These norms are not applicable to the larger model. (Kent and Shakow, '28, p. 120)
Discussion: These boards represent a high degree of experimental procedure and refinement in test construction, being the final result of several years' research. According to their makers, "The test places a high premium upon careful planning as opposed to trial-and-error performance." The factor of chance has been almost entirely eliminated, as each problem is presented five times, thus acting as a check on any chance placements which might occur. Practice effects can be studied without requiring the Subject to perform the same task over and over again. Other than intellectual individual differences can be detected, as the increasing difficulty of the problems stimulates some people, whereas it discourages others. (Kent and Shakow, '28, p. 119)
References:
Shakow and Kent, '25
Kent and Shakow, '28

No. 117 Lincoln Hollow Square

Devised by E. A. Lincoln.
Date: 1923
Handled by Psycho-Educational Clinic, Palfrey House, Harvard; also by Stoelting.
Price: $3.50. Manual, with other data; seventy-five cents. (Palfrey House); $7.00 (Stoelting).

*Dr. Kent can be reached at Box 50, Hathorne, Mass.

Material: Board $7\frac{1}{2}$ x 10 x $\frac{5}{8}$ inches, into which is cut a hole $4\frac{1}{2}$ inches square; eleven blocks, eight of different shapes. With these pieces different combinations may be made which will exactly fill up the hole.



Reproduced by courtesy of C. H. Stoelting Co.

Problem: To fill up the hole as quickly as possible.

Individual test; eight problems to be solved. This test is not adaptable to group testing.

Time: Twelve to fifteen minutes. On minute limit for each problem.

Scoring: Total time for the eight problems.

Standardization: This test was standardized by Lincoln on 500 cases, based on three series of distributions; hospital cases, clinic cases and private school children. Lincoln's standardization is based on: 1, total time scores by chronological age for all cases; 2, total time scores by chronological age for all cases between 90 and 109 I. Q. 3, total time scores by mental age. The assumption is that the performance of the normal group will give the same average as the group as a whole, and that the distributions for any mental age will give the same central tendency as the distribution of the corresponding chronological ages. ('27, p. 265f) Norms are expressed for time in terms of age. Range: from two to eight years, although the norms go to ten years.

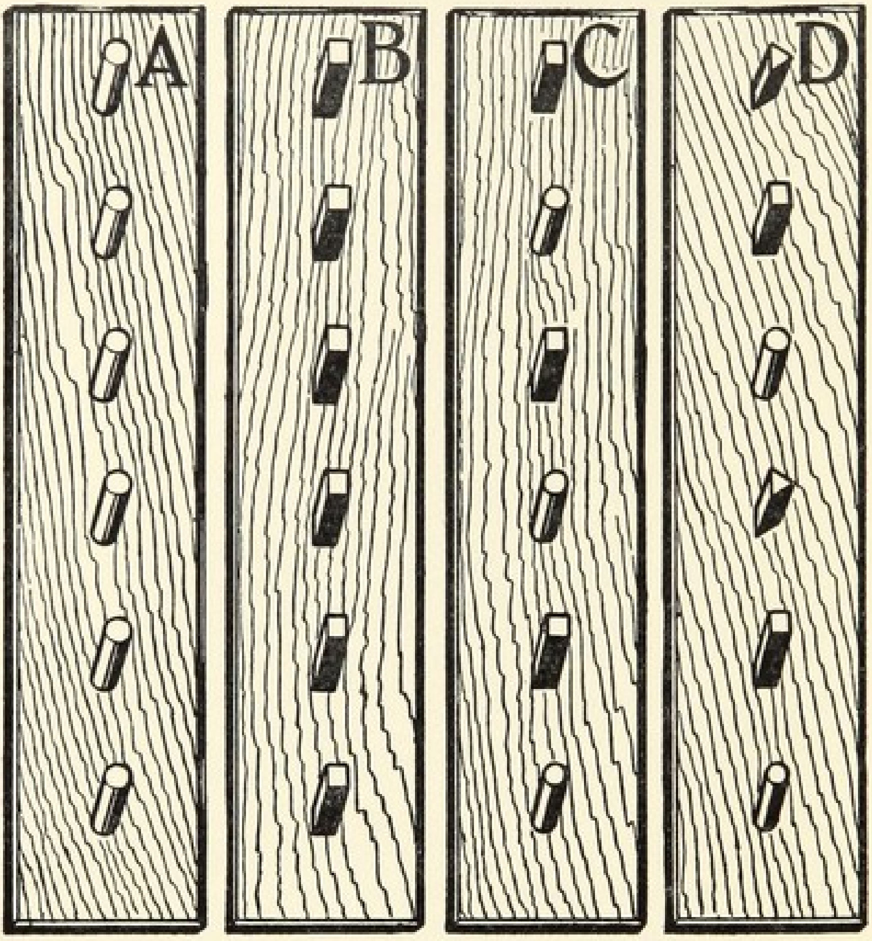
Reliability: $r = .95$, N being 139.

Discussion: This, according to its author, is a test of form concepts and manipulative skill. F. L. Wells considers it a formal improvement over the construction puzzle tests, in that it provides solutions of graded difficulty for the same cut-out space. ('27, p. 136)

References:

Dearborn, Shaw, Lincoln, '23
Lincoln, '27
Bronner, Healy, Lowe, Shimberg, '27
F. L. Wells, '27

No. 118 Peg Boards, A. B. C. D.



Reproduced by courtesy of C. H. Stoelting Co.

Devised by J. E. W. Wallin.
Date: 1918
Handled by Stoelting.
Price: $12.25

Material: Four boards, each 3 x 14 x ¾ inches. Each board contains recesses for six pegs, similar in size but different in shape. Board A has round pegs only, B has square pegs only, C has three round and three square pegs alternately arranged, and D has two round, two square and two triangular pegs, alternately arranged.

Problem: To put the pegs in their appropriate holes.

Individual test: four different tasks, varying with the boards. This test is not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Boards A and B: one hundred seconds each. C. and D: sixty seconds each.

Scoring: Time on Board A, plus time on Board B, plus one-half time on C and D, and one-third time on D. (Goodenough '25, p. 206)

Standardization: This test has been standardized by Goodenough on 300 cases of two- three- and four-year-olds; (Goodenough '27) also by Hallowell on

614 children, ages one to four. (Hallowell '28) Norms are expressed for time in terms of half-year ages, sex and deciles. Range: one and one-half to four and one-half years. The easier boards can be extended downward as far as twelve months.

Reliability: By re-test after six weeks. r = .794, age two. r = .688, age three. r = .583, age four.

Validity: r = .507 with the Kuhlmann-Binet (Goodenough '27, p. 215)

Discussion: Hallowell thinks that the Wallin Peg Boards demand more highly developed motor co-ordination in young children than does the Three Disc Formboard. Series C and D require analytic discrimination. Since the less able children work trial-and-error wise, it becomes a test of persistence as well as of form discrimination. Hallowell thinks that the test is not diagnostic above the age of three. Below age two, only Series A and B are applicable. (Hallowell, '28, p. 254 f) Goodenough found that, although the children from the upper socio-economic levels did better on the Kuhlmann-Binet test than did children from lower levels, there were no marked differences in their ability to handle the problems presented by the Peg Boards. She adds, "It may be that the peg board performance is in some degree related to mechanical aptitude, the Binet to potential scholastic achievement. This hypothesis is plausible, but it has not been verified." (Goodenough '27, p. 214 f)

References:

Wallin, '18
Baldwin and Stecher, '25
Stutsman, '26
Goodenough, '27
Hallowell, '28
Baldwin, '28

No. 119 Seguin Form Board

Devised by Seguin; modified by N. Norsworthy and H. H. Goddard.

Date: First used as a test in 1906.

Handled by Stoelting; also by Marietta.

Price: $12.50 (Stoelting). $6.50 (Marietta).

Material: Board, 13.5 x 18.5 inches with ten recesses; ten wooden blocks. Stoelting lists a revision by Goddard and Sylvester as an "inverted form" of the above; (price $15.00). Stutsman used a "flush" model ($10.25), and Whipple's modification, (price $17.00) introduces a "series of removable blocks, the position of which may be changed at will. The blocks fit into a hinged case which, when closed, makes an excellent carrying case" (Stoelting 1930 Catalogue). Drever and Collins also have produced a variation of this test.

Problem: To fill the recesses with the appropriate blocks.

Individual test; three trials. This test is not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Time limit of five minutes for each of three trials.

Scoring: Time on each trial. Shortest time is scored.

Standardization: Norms are expressed for time in terms of age, sex, (Wallin) and in percentiles. This board was standardized by Wallin on 1345 cases, ranging in age from five to fourteen years.

Discussion: See Seguin Formboard (Witmer-Sylvester modification). No. 120 Besides testing intelligence (Witmer) this board has been said to test perception, discrimination of forms, manual or constructive ability, motor co-ordination, ability to meet new situations, planfulness and learning ability, and so fourth. These claims are listed but not evaluated by Bronner, Healy, *et al.* ('27, p. 221 f)

The procedure for administering a test such as this, can be varied. For example: the Subject may be required to place the blocks in the recesses, first with vision; and then tactually, while blindfolded. Learning under this second condition, and general rate of improvement can also be studied.

References:

Norsworthy, '06
Goddard, '12b
Wallin, '12, '16, '21, '27
Whipple, '14
N. Y. Bureau of Analysis and Investigation, '15
Gaw, '25
Baldwin and Stecher, '25
Stutsman, '26
Bronner, Healy, Lowe and Shimberg, '27
Arthur, '28
Drever and Collins, '28

No. 120 Seguin Form Board

Devised by Seguin; modified by L. Witmer and R. H. Sylvester.

Date: 1911

Handled by Psychological Laboratory and Clinic, University of Pennsylvania: also by Stoelting.

Price: $25.00 (Stoelting).

Material: Form board, 20 x 14 x ⅜ inches, and eleven geometrical pieces. Tray for holding pieces. The "Witmer" board is smaller than the "Goddard" board.

Problem: To put the blocks in the board as quickly as possible, blocks to be arranged according to directions; three trials.

Individual test. This test is not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Scoring: 1, time of first trial in seconds; or 2, time of shortest trial in seconds.

Standardization: This test was standardized by H. H. Young on 1793 males and 1691 females, 845 males and 802 females (Young '16) Norms are expressed for time in terms of age, sex and percentiles. Range: five years to adult.

Discussion: Witmer considers this to be one of the best tests to distinguish feebleminded children from those of normal intelligence. (Witmer '11) Young has prepared a list of forty-nine functions, mental and physical, which can be observed while the subject is using the board. (Young '16) Bronner, Healy, *et al*, sum up the offerings of other workers but point out that the interpretation of test results is dependent upon test conditions and test procedure. (Bronner, Healy, Lowe and Shimberg, '27, p. 221 f) Kent considers the "Witmer Formboard", as it is often referred to, as being one of the very best for the primary stage. (Personal interview) Hallowell also claims that the Witmer Formboard is especially valuable for pre-school children, many of whom do not get the idea of trying the block in another hole, and are content to have the block fit the wrong hole. ('28, p. 252) F. L. Wells suggests that for earlier years "perception of form" is the governing factor, and that for older people it becomes a matter of manual skill. Failure to differentiate the form may be indicative of a transitory lapse of attention as well as of poor intelligence. The interpretation, if not the scoring, is thus affected. (F. L. Wells '27, p. 132).

References:

Witmer, '11
Sylvester, '13
H. H. Young, '16a and b
Pintner and Paterson, '16a, '17
Ide, '18
H. H. Young and M. H. Young, '23
Bronner, Healy, Lowe, Shimberg, '27
F. L. Wells, '27
Hallowell, '28.

No. 121 Three Disc Form Board

Devised by L. Witmer.

Date: 1928.

Handled by the Psychological Clinic Laboratory, University of Pennsylvania.

Price: $5.00.

Material: Wooden board, 8¾ x 4¼ inches, with three recesses ¼ inch deep and ¾ inches apart, into which fit round wooden blocks, 1¾ inches in diameter and 7/16 inches thick.

Problem: Two trials to put the blocks into the board. If necessary ten teaching trials may be allowed.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Time element not considered except as: rapid, slow, and so forth.

Scoring method: Success, learning or DNC (did not complete), or failure.

Standardization: Hallowell has standardized this test on 199 pre-school children. Norms are expressed in terms of age and deciles. Range: one year to two years.

Reliability: This test correlates well with other tests of similar type. (Hallowell letter).

Validity: No correlation statistics have been offered.

Discussion: Hallowell found that this board could pick out the superior children at the age of sixteen months, at which age only nine per cent of the children will succeed. At eighteen months, more than fifty per cent will succeed, thus making the test less diagnostic of superior ability at this and upper ages. However, she stresses that no individual test at these low ages is to be considered by itself as significant for intelligence testing purposes. It should be included and evaluated in the light of a range of responses to other test situations. She also used the board to test learning ability, by studying the number of repetitions needed to solve the test without help. (Hallowell '28, p. 247).

References:

Hallowell, '28

No. 122 Three Figure Form Board

Devised by L. Witmer.

Date: 1928

Handled by the Psychological Laboratory and Clinic of the University of Pennsylvania.

Price: $5.00

Material: Wooden board, 8¾ x 4¼ inches, with three recesses and three blocks shaped in order: (a) a circle, (b) an equilaterial triangle, and (c) a square.

Problem: To put the blocks in the board; three correct trials, one with board in reversed position. • Ten teaching trials allowed.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Time element not considered.

Scoring method: Two correct trials in position A plus next immediate one in B; or *vice-versa,* two in position B and one in A.

Standardization: Hallowell has standardized this test on 594 pre-school children, and offers a decile distribution. Range: one year to three years, eleven months. For three-year-olds, the test has little value, except as introductory to the Witmer form board. Below eighteen months, all responses were failures.

Validity: This test correlates satisfactorily with other tests of equal difficulty.

Discussion: Hallowell found that this test was excellent to study the early appearance of form discrimination and motor co-ordination in the child. Insertion of the square and the triangle in this board requires more highly developed motor co-ordination than do the circles of the Three Disc Form Board. She noted also that the brighter children were the first to give up trying to force a square block into a round hole. Persistence in such an attempt was characteristic of the very young or the duller children. (Hallowell '28, p. 249)

References:

Hallowell, '28

No. 123 Two Figure Board

Devised by R. Pintner.

Date: 1917

Handled by Stoelting; also by Marietta.

Price: $5.00 (Stoelting): $4.50 (Marietta).

Material: Wooden board, 38.3 x 25.4 x 1.4 centimeters, with two cut-outs, a square and a cross; and nine pieces: square, triangles and rectangles.

Problem: To fit the pieces into the board as quickly as possible.

Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time and moves.
Standardization: Pintner and Paterson standardized this test on 911 school children. Norms are expressed for time and moves in terms of age, percentile and quartiles. "The distribution suggests a slight element of chance due to the puzzle nature of the board. The scattering at some ages is very great." Range: five to fourteen years. (Pintner and Paterson '17, p. 108)
Discussion: Bronner, Healy, *et al*, report that the Two Figure Board is easier than the Five Figure Board, and point out that there is considerable irregularity above nine years of age. ('27, p. 219) Arthur found that this was the least discriminative test in her scale. ('25, p. 392–395) Drever and Collins have produced a modification of this test.
References:
Pintner and Paterson, '17
Arthur, '25, '28
Johnson, '25
Bronner, Healy, Lowe and Shimberg, '27
Drever and Collins, '28

No. 124 Wiggly Block Test

(Work sample No. 5. General Electric Co.)

Devised by J. O'Connor.
Date: 1920
Manufactured by the General Electric Co., Lynn, Mass. Available only to recognized testing laboratories intending to use this test for research in mental testing.
Material: A block of wood, 10 x 6 x 6 inches, jig-sawed into nine sections longitudinally, in a "wave-like" pattern.
Problem: To put together the scattered parts in the quickest possible time. The test is given three times consecutively to one person.
Individual test; not adapable to group testing.
Directions: Part verbal, part pantomime.
Responses: Manipulative.
Time: Fifteen to twenty minutes.
Scoring: Time, taken with a decimal timer stop-watch. Appropriate factors are used to reduce each score after the first trial to an equivalent first-trial score. Average of three trials.
Standardization: Norms are expressed in time, and are arranged in quartiles, for males and for females. Factors are used in the scoring for ages under twenty. O'Connor has standardized the tests on about 10,000 cases. (letter to the writers)
Discussion: This test measures the ability to visualize three-dimensional structure, an essential and inherent aptitude for success in all types of engineering. It is a test of mental endowment for imagining concrete space. (letter)
References: Keane and O'Connor, '27 O'Connor '28, and letter to writers

No. 125 Worcester Form Boards

Devised by D. C. Shakow and G. H. Kent.

Date: 1925

Handled by Stoelting; also by Ellen Wilson, Rome, N. Y.

Price: $21.75 for first two boards in the set. (Stoelting). $38.70 for the four boards.

Material: Four form boards, made of wood, size 26 x 20 centimeters; each with a number of irregular cut-outs. The third board has four sets of blocks and the fourth board has two sets.

Problem: Blocks arranged according to directions. Subject is required to put in the blocks where they belong as quickly as possible.

Individual test.

Scoring: The scoring method has been recently revised by the test authors. (Kent, letter)

Discussion: This test has been discarded by its authors, and is superceded by the more recent "Graded Series of Form Boards", Test No. 116.

References:

Shakow and Kent, '25

Bronner, Healy, Lowe and Shimberg, '27

Kent, letter to writers

X

MAZES

No. 126 High Relief Finger Maze

Devised by W. R. Miles.
Date: 1928
Handled by Stoelting.
Price: $27.50.
Material: Wooden base into which are inserted raised paths formed by wire staples; small practice section in lower right hand corner.
Problem: Subject, blindfolded, after preliminary practice, tries to learn the maze pattern by finger tracing.
Individual experiment; not adaptable to group work.
Directions: Verbal; pantomime doubtful.
Responses: Manipulative.
Time: No limit.
Scoring: Time; errors.
Standardization: This maze is not standardized as an intelligence test, but was tried out experimentally by Husband who used eighty subjects to study learning ability. He gives a coefficient of correlation "between trials to learn and errors" of $.84 \pm .02$. (Husband '28, p. 23)
Discussion: In general, learning is much facilitated by using a finger maze form, and therefore more suitable for children. The general advantage of this type of maze is that the Subject has direct cutaneous contact with the pattern, does not have the experience of pressing on the wrong side as with pencil or groove mazes, and he can traverse the maze more rapidly. The wire maze is less expensive than the slot type. (Miles '28, p. 13)

Husband gives as factors in learning: "Verbal or counting; forming a visual pattern; a combination of verbal, visual and motor; verbal with some other method discarded before completion of learning; verbal with some other method employed secondarily, and motor last." (Husband '28, p. 28)

References:
Miles, '28
Husband, '28

No. 127 Maze A and B

Devised by W. S. Foster.
Date: 1923
Handled by Stoelting.
Price: $8.75 each.

Material: Fiber board on which is a "grooved" maze pattern: stylus or lead pencil.
Problem: Blindfolded Subject must work his way out to the goal in the shortest possible time.
Individual experiment; not adaptable to group work.
Directions: Verbal; pantomime with difficulty.
Responses: Non-verbal.
Time: No limit.
Scoring: 1, time on repeated trials to get learning curve; 2, errors; 3, combination of two.
Standardization: This maze has not been standardized as a test of intelligence, but is included here because of its significance as a test of learning and habit formation. Foster gives average scores made by a group of fifty-eight and sixty-five subjects on Maze A and B respectively, for each of twenty-five trials. (p. 162)
References:
Foster, 23 (p. 158 ff)
Foster and Tinker, '29 (p. 181 ff)

No. 128 Maze Tests

Devised by M. T. Whitley.
Date: 1911
Handled by Stoelting.
Price: $0.80 per hundred for each pattern.
Material: Five different maze patterns: curved, straight, combined, black and spiral; drawn on paper.
Problem: To draw a line between the two lines without touching either.
Individual or group test; five parts.
Directions: Verbal; pantomime by adaptation.
Scoring: Number of seconds plus five seconds per touch. (Whitley '11, p. 90)
Responses: Non-verbal.
Standardization: No norms are offered for these maze tests, but Whitley reports using them in her investigation of tests of individual differences. One of these forms, the spiral, was used by Simpson on his group of thirty-seven adults: seventeen superior and twenty inferior. He refers to it as the "Scroll" test. No standardization is offered, but the following data were obtained.
Reliability: r = .76 for "Scroll" test only; first and second trials.
Simpson states that this reliability can be raised by giving opportunity for preliminary practice. (Simpson '12, p. 43)
Validity: r = .26 for "Scroll" test against the other eleven tests in Simpson's battery. (Simpson '12, p. 66)
Discussion: In explaining why the coefficient of reliability is higher for the poor group than for the good group, Simpson notes that the individual differences brought out in the poor but not in the good group are probably in part due to the difference in facility in handling pen and pencil; the bad performance of some in the poor group is attributed to weak muscular control

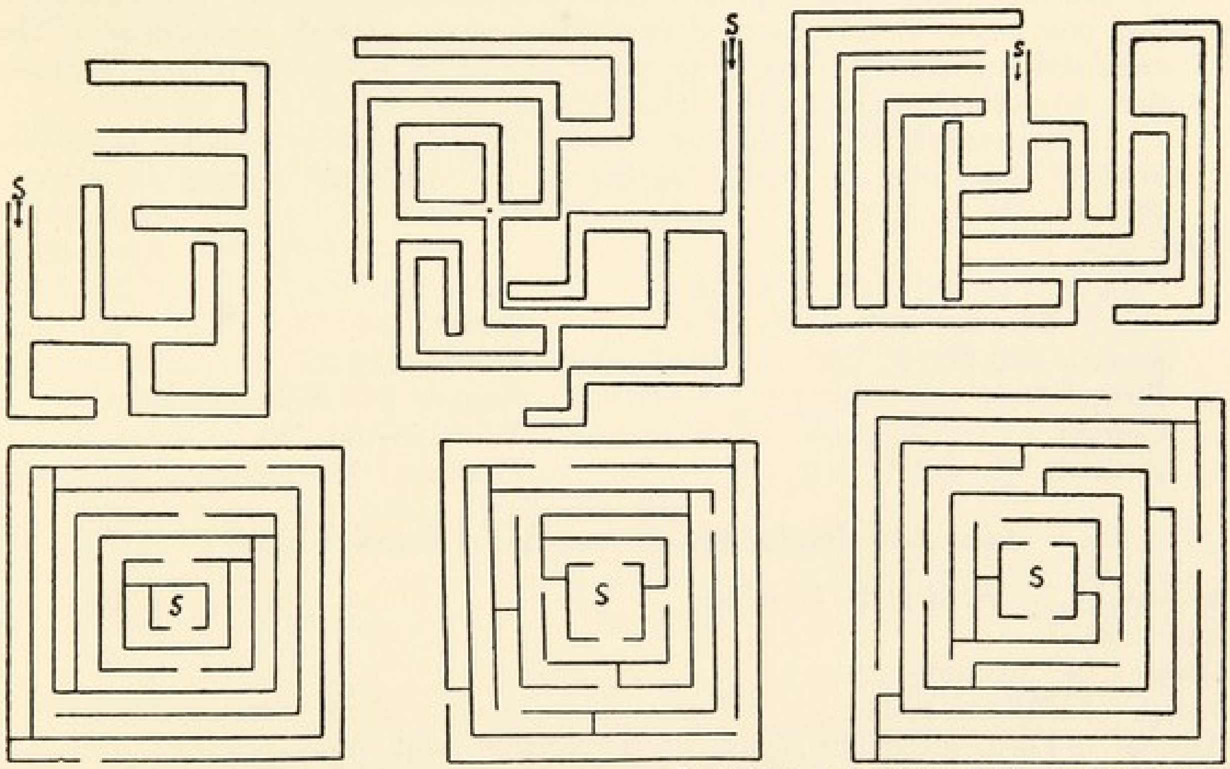
and unsteadiness of hand. The Scroll test is also somewhat vitiated in that it tends to cause dizziness and eyestrain.

References:

Whitley, '11 Simpson, '12

No. 129 Porteus Mazes

Devised by S. D. Porteus.
Date: 1915; date of revisions: 1916, 1919
Handled by The Training School, Vineland, N.J.; also by Stoelting.



Reproduced by courtesy of C. H. Stoelting Co.

Price: $0.80 per pad (Stoelting). Price of manual: $1.15
Material: One hundred blanks on a pad; one pad for each age. Examiner detaches each test blank when used. 1100 test blanks weigh 3.2 pounds.
Problem: To trace with a pencil the shortest way out of each maze.
Scale: Eleven sets of mazes; two adult mazes.
Individual or group test. Character of test is changed if used as a group test.
Directions: Verbal.
Responses: Non-verbal.
Time: Fifteen to twenty minutes.
Scoring: Year scale with deductions from best age for errors.
Standardization: The tests have been standardized by their author, (adults by Porteus and Bassett), on 2488 cases including 1000 Australian children, 1255 Australian and American children, and 663 adults. Norms are expressed in terms of age. The tests are specially applicable to pre-school children, primitive peoples, socially maladjusted, defectives and borderline cases.

Reliability: r = .95.

Validity: r = .70. Criterion: social rating obtained in institutions, and observed success on parole, and so forth.

Discussion: According to their author, these tests purport to measure the tendency to use forethought and prudence in a concrete problem; also a complex of other mental and temperamental traits. They could be used for testing peoples of different cultural background; in fact these tests have been most widely used for this purpose. Porteus states, however, that his recent experience in Australia has led him to believe that all tests are affected by the racial background. In other words, no test is free from the effects of training. In Porteus' opinion, there is no need for tests of this sort, since "no one in his development is free from environmental experiences". The maze test, however, is freer from the effects of inequality as regards training than any other test he knows. (letter to writers)

"In consideration of the fact that the Porteus scale is founded upon one kind of test only, it would appear to be distinctly limited in scope." (Squires '26, p. 3f.)

References:

Porteus '15a and '15b, '17, '18, '19, '22, '24a and '24b, '25
Porteus and Berry, '20
Burt, '23
Cornell, E. L., '24
Arthur, '25
Gaw, '25
Jarrett, '26
Squires, '26
Porteus and Babcock, '26
Worthington, '26
Poull and Montgomery, '29

No. 130 Right Hand and Left Hand Mazes

Devised by L. W. and F. L. Kline.

Date: 1927

Handled by Stoelting.

Price: $32.50 per maze.

Material: These mazes are 30 x 30 centimeters, with one-centimeter runways. They are made of five-ply wood, mounted on a base, with a screen at the top and three sides; curtain at rear to permit insertion of arm; stylus and rubber finger guide; sheets of paper, about 32 x 32 centimeters; stop watch.

Problem:

1. Subject inserts right arm through the open side. His problem is to learn the maze pattern by moving the stylus along the runways until the goal is reached. 2. To do the same with left hand.

Individual experiment; too complicated for group use.

Directions: Verbal; non-verbal by adaptation.

Responses: Non-verbal.

Time: No set limit.

Scoring: The maze is learned when three successive tracings have been made without error or stopping.

Standardization: This material has not been standardized as a test of intelligence, but is included here because of its possibilities for experiments and tests on learning.

Discussion: Material of this sort can be used to advantage in studying the rate of learning and adaptation, unlearning or forgetting, re-adaptation to a slightly altered situation, and many other variations of the learning situation.

References:

L. W. and F. L. Kline '27

No. 131 Slot Maze A

Devised by H. H. Young.
Date: 1922
Handled by Stoelting.
Price: \$22.50
Material: Board with maze pattern cut out in metal, equipped with a toy shoe stylus and the figure of a boy at the goal; self correcting.
Problem: To take the shoe to the boy as quickly as possible.
Individual test; two trials. This test is not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time limit: Five minutes.
Scoring: Time; two trials allowed.
Standardization: This test was standardized by H. H. Young and D. C. Daugherty on 3141 children. Norms are reported for age and sex and distributed in quartiles; two trials separately. Range: four to eleven years.
Reliability: $r = .618 \pm .017$ on crude time scores, trials five to fifty, odd versus even, on normal three-year-olds.
Validity: "Ability to see a maze pattern (initial trial) and know it (total time and errors) is not related very highly to the function which is measured by scores on an intelligence test." The range of Rho was from —.20 to + .68. (McGinnis '29, p. 254)
Discussion: "The distinct sex difference in favor of the boys at each age and in all three experiments (Young's and Easby-Grave's) are also evident." .... "The large initial sex differences tend to disappear when learning is continued for fifty trials." (McGinnis '29, p. 249)

References:

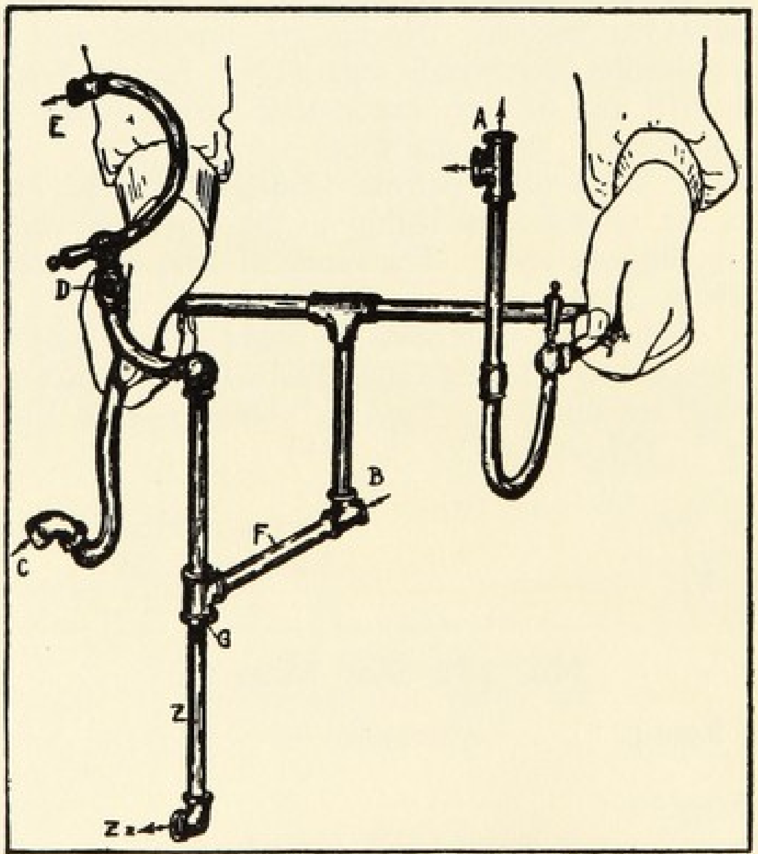
H. H. Young, '22
Easby-Grave, '24
A. M. Jones, '25
Bronner, Healy, Lowe, Shimberg, '27
McGinnis, '29

No. 132 Slot Maze C

Devised by H. H. Young.
Date: 1930.
Handled by Stoelting.
Material: Board with maze pattern cut out of metal, equipped with a stylus which cannot be removed except at start and goal.
Problem: To take the stylus from the starting point to the goal as quickly as possible.
Individual test; two to five trials.

Directions: Verbal.
Responses: Manipulative.
Time limit: Ten minutes.
Scoring: Time.
Standardization: This test is being standardized by Young. Range: ten years through adulthood.
Discussion: A. M. Jones, in her study of 120 superior children, found a negligible relation between success on the maze tests and general competency. She thinks that the ability to solve Slot Maze C is something specific, although she admits a conflict of opinion on this point among the other workers in the University of Pennsylvania laboratory. However, she concludes by remarking that the high percentage of failures on this test, as contrasted with success on the Dearborn Formboard, would suggest a different set of abilities. Too much imagination on the part of the child leads him to suspect a trap, and thus tends to lower his chances of quick solution. (A. M. Jones, '25, p. 51)
References:
A. M. Jones, '25. H. H. Young. (letter to writers)

No. 133 Tri-dimensional Maze



Reproduced by courtesy of C. H. Stoelting Co.

Devised by D. P. Boder.
Date: 1930
Handled by Stoelting.

Price: $16.50

Material: Small 3/8-inch brass pipes fastened together; holes to act as *cul-de-sacs;* adjustable for difficulty.

Problem: To roll around a 3-16-inch steel ball-bearing in such a manner that it emerges at a previously designated goal. Problems may be arranged for varying degrees of difficulty.

Individual experiment; not adaptable to group work.

Directions: Verbal; pantomime possibly.

Responses: Manipulative.

Time: No limit.

Scoring: 1. Time. 2. Number of attempts. 3. Combination of the two.

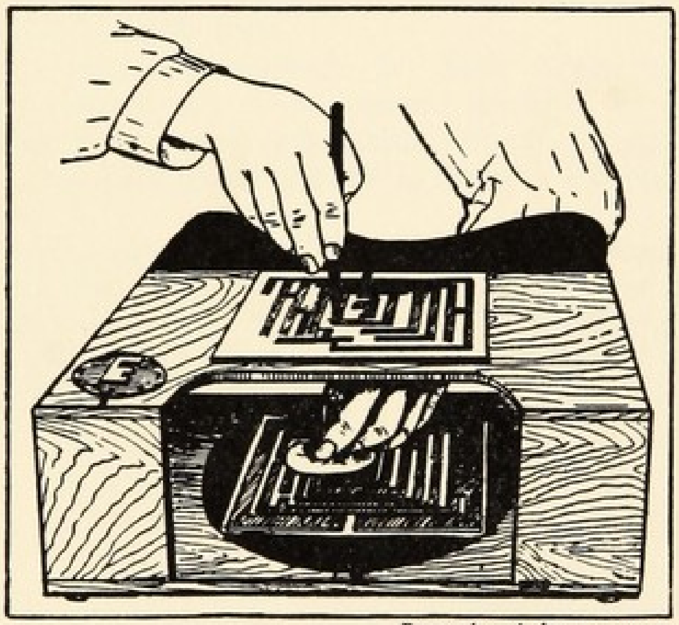
Standardization: This maze has not been standardized as a test, but because it offers promise of indefinite variation, is included here.

Discussion: Tasks from very simple to very complex can be arranged, depending on the goal set in the maze, each requiring more and more delicate movements and co-ordinations on the part of the Subject.

References:

Boder, '30

No. 134 Two-Story Duplicate Maze



Reproduced by courtesy of C. H. Stoelting Co.

Devised by W. R. Miles.

Date: 1919

Handled by Stoelting.

Price: $55.00

Material: Maze in a covered space below corresponds to a visible maze above. Maze below is rotated at different degrees, and then inverted, thus creating different problems.

Problem: Subject must find his way through the lower maze without the use of direct sight.

Individual experiment; not adaptable to group work.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: One to fifteen minutes according to difficulty of problem set.
Scoring: Time to solve problem.
Standardization: This test was experimentally tried out by C. M. Cox on sixty-seven college students. Its author, Miles, says it is adaptable to primitive groups and superior adults alike.
Discussion: This test involves "the ability to understand and manipulate space relations which are out of sight but fully represented in sight." It is "also subject to training." (Miles, letter to writers).
References:
Miles, '27, and letter to writers.
Cox, '28a and '28b

XI

TESTS INVOLVING RECOGNITION OF FORM, COLOR, WEIGHT

No. 135 (A) Bilaterally Symmetrical Tests (B) Radially Symmetrical Tests

Devised by T. H. Briggs; modified by F. W. Steacy.
Date: 1913
Material: Each form has a page of fifteen figures, eight of which are symmetrical.
Problem: To mark a line through the figures that are (A) bilaterally and (B) radially symmetrical.
Individual or group test. Two forms.
Directions: Verbal; pantomime with difficulty.
Responses: Non-verbal.
Time: Three minutes for each form.
Scoring: Each figure correctly marked: 2. Each figure incorrectly marked: minus 2 (in star: minus 1). Unattempted: minus 8.
Standardization: This test is not standardized, although Briggs and Steacy report working with it on sixth grade groups in New York public schools.
Reliability: r = .04.
Validity: r = .34 with other tests in Steacy's battery. (Steacy, '19, p. 42)
Discussion: This test involves understanding the meaning of the terms "bilaterally and radially symmetrical" and ability "to thoroughly apply a definition." (Briggs, '13, p. 22) Also "recognition of similarities and differences in form are essential requirements." (Steacy, '19, p. 42)
References:
Briggs, '13
Steacy, '19

No. 136 Block Designs

Devised by F. N. Maxfield; modified by S. C. Kohs.
Date: 1919
Manufactured by Stoelting.
Price: $3.25. Scoring tables: $0.85. Scoring blanks: $2.50 per hundred.
Material: A series of seventeen designs, increasing in difficulty. Box of sixteen colored cubes, approximately one inch in size. The cubes each have red, blue, yellow and white surfaces. Each remaining side is divided diagonally into a blue-yellow and a red-white combination of colors. Drever and Collins offer a modification of the Kohs test, which requires only ten of the original seventeen designs.

Problem: To reproduce the designs according to the pattern.

Individual test: This test is adaptable to group testing, but with difficulty.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: About one hour.

Scoring: Time, moves and accomplishment.

Standardization: This test has been standardized by Kohs on about 400 normal boys and girls. Norms for time, moves, and score points for each design, are expressed in terms of age. Range: five to twenty years. The test is applicable to all kinds and types of people. (Kohs, letter to writers)

Reliability: "The block design test appears to have about 56 per cent the diagnostic power of the Binet vocabulary." (Kohs, '23, p. 208)

Validity: $r = .82 \pm .01$ with Stanford-Binet age; 366 cases. (Kohs, '23, p. 181)
$r = .71 \pm .02$ with Army test; 167 cases. (p. 202)
$r = .57 \pm .03$ Trabue, B and C.; 273 cases. (p. 201)
$r = .46 \pm .04$ with Nat. Int. Test; 104 cases; nine, ten and eleven-year-olds. (A. J. Brown, '30)

Discussion: Blackwood objects to the use of this test for inter-racial testing on the ground that it would give an advantage to the kindergarten child who has become familiar with it through play materials. ('27, p. 54) Kent in giving this test to superior adults found that they all made high scores but that the number of moves depended more on temperament than on the mental ability measured by the test. She noted that some toyed aimlessly, others worked quickly, trial-and-errorwise, while still others made accurate and deliberate motions. ('23) This test, according to its author, measures intelligence. Wells, also considers it sound for that purpose. ('27, p. 140)

Brown who analyzed the results obtained on 154 boys and girls, aged nine, ten and eleven years, found arithmetical means which closely approximated Kohs' medians. "But," she adds, "when these averages are interpreted in the light of their arithmetical deviations, the validity of the test is called into question." Although she recognizes that Kohs found an r of $82 \pm .01$ with Binet mental ages, she concludes that "judged by the principle of demarcation of age groups, however, the validity of the Kohs test is of doubtful value." (Brown, '30, p. 181)

References:

Kohs, '20 and '23, and letter to writers.
Kent, '23
Maxfield, '25
F. L. Wells, '27
Drever and Collins, '28
"Proceedings" (Reymert), '30
A. J. Brown, '30

No. 137 Color Cube Test

Devised by F. N. Maxfield; modified by R. B. W. Hutt.

Date: 1925

Manufactured by Embossing Co.; also by Stoelting.

Price: $0.35 (Embossing Co.); $6.25 (Stoelting)

Material: Box of blocks; 4¼ x 4¼ x 1¼ inches in bulk, consisting of eight color cubes, similar to those in Kohs' test; a set of five designs.

Problem: Given practice in copying from a model, to reproduce a design after the pattern is removed. (memory)

Individual test: Five items, A, B, C, D, E; not readily adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Ten seconds exposure for D; five seconds for E.

Scoring: Time. Scores are based on success or failure in three minutes. Form of incorrect solutions recorded by means of symbols or sketches.

Standardization: This test was standardized by Hutt on 1198 normal children; norms are expressed in terms of age, grade, and in percentiles. Range: six to fourteen years. It is applicable to normal people from six to twelve years, and to mentally defective adults of imbecile or moron levels.

Discussion: The first three designs are reported to be too easy for diagnostic purposes above six years. The test is useful to test "imageability," involved in perception and memory; also factors of attention, perception, memory play a large part in the solution. Planfulness, observation and intelligence show forth; also trainability, if test is repeated. The range of factors involved must serve as a warning that account must be taken of all in some other way, if imageability is to be tested. Differences in time reactions seem due to factors other than imageability. (Hutt, '25, p. 95f)

References:

Town, '21
Kohs, '23
Hutt, 25
Fernberger, '27

No. 138 Color Cube Test

Designed by F. N. Maxfield.

Date: 1925

Manufactured by Embossing Co.; also by Stoelting.

Price: $0.35 (Embossing Co.); $6.25 (Stoelting).

Material: Box of blocks 4¼ x 4¼ x 1¼ inches in bulk; including eight color cubes and a set of five designs. (Same as Hutt's. See Test No. 137)

Problem: With the color cubes to reproduce patterns from a design set in front of the Subject. (Imitation or learning).

Individual test, consisting of a series of five designs of increasing difficulty, A. B. C. D. E. This test is adaptable to testing very small groups.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Seven to ten minutes.

Scoring: Time and degree of success or failure in the difficulty scale. Close attention should be paid to the qualitative aspects of the performance.

Standardization: This test is not standardized except for order of difficulty of the designs. It is suitable for children from four to ten years, for feeble-minded adults, and special-class adolescents.

Discussion: The advantages of this test are that the Examiner is sure the test is understood; the language element is reduced to a minimum; shyness and self-consciousness are readily overcome; interest is aroused. The test comes as a graded series with a fairly wide range of difficulty. The test can be repeated, using different designs but equal in difficulty. The apparatus is simple, inexpensive and easily portable. The test can be given to two or three Subjects at the same time. (p. 102 f)

"The examiner must take into account the planfulness of the hit-or-miss method of the boy, his motor co-ordination, his quickness in noting which block requires adjustment, his persistence and his energy. In this opportunity for observing the boy's 'mind at work', lies the chief value of the test." (Maxfield, '25, p. 101) Maxfield says also: "I find this test useful because it is *not* standardized and therefore is a very flexible instrument in clinical use as a supplement to the more rigid situation of a standardized procedure . . . . I change to the memory test if I find the other too easy, but keep the same designs. . . . One can repeat in new color combinations: A—3 times or 4 in all; B—5 times or 6 in all; C—3 times or 4 in all; D—3 times or 4 in all; E—3 times or 4 in all; and still get the motivation of trying a new task. D is rather difficult for most feeble-minded adults." (letter to the writers)

References:

Maxfield, '25 (illus.), and letter to writers
Hutt, '25

No. 139 Color Cube Test

Devised by S. C. Kohs; modified by G. H. Kent.

Date: 1927

Manufactured by Stoelting.

Price: $20.00 (Stoelting); or blocks which are on sale at department stores can be bought for about 35 cents. (Ask for Embossing Co's Color Cubes); or designs can be painted at home.

Material: Sixteen color cubes, twelve designs.

Problem: To arrange blocks according to a design placed in front of the Subject.

Individual test: Twelve tasks, graded in difficulty. This test is adaptable to group testing with difficulty.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Twenty minutes.

Scoring: Time converted into mental ratings according to a table.

Standardization: This test has been standardized by Kent, assisted by Shakow and others, on about 300 children. Norms are expressed for time in terms of age. Range: from six years to superior adult. Kent states that norms given by Bronner, Healy, Lowe and Shimberg, ('27, p. 148) are no longer in use. She herself has prepared revised norms.*

*Dr. Kent can be reached at Box 50, Hathorne, Mass.

Discussion: Kent has enlarged the original designs so as to make them the same size as the blocks, because non-English speaking subjects do not grasp the connection between blocks and designs unless both are the same size. She has found that the blocks appeal to girls more than to boys: Kent does not think this test could be used on people of different racial cultures unless especially standardized for them. (letter to writers)

References:

Bronner, Healy, Lowe, Shimberg, '27 Kent, letter to writers

No. 140 Color, Form and Size Discrimination Test

Devised by B. M. Boody.

Date: 1926

Material: Sixty small forms in three sizes, about one inch, one and a half inches and two inches in diameter, comprising squares, triangles, diamonds, circles, octagons; cut out of red, green, yellow and blue cardboard. Each piece was numbered on the back. A duplicate set of ten pieces for Examiner's use. Into each of these ten samples a pin is set for a handle. Examiner's set consists of:

1. Square; red; medium
2. Triangle; yellow; small
3. Diamond; blue; large
4. Circle; green; large
5. Octagon; red: small
6. Square; blue; small
7. Diamond; green; medium
8. Octagon; red; large
9. Triangle; green; small
10. Circle; yellow; medium (p. 122f)

Problem: Experimenter holds up sample which is always presented in fixed order from one to ten; child selects from among the sixty pieces the one which corresponds to Examiner's sample in color, size and form; ten preliminary practice trials allowed.

Individual test; not readily adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: Recorded but not counted in scoring. Range from 82 to 138 seconds.

Scoring: Number correct.

Standardization: While this test cannot be said to have been standardized in the strict sense of the word, nevertheless Boody offers results obtained on sixty immigrant boys and girls at Ellis Island, ranging in age from eight to sixteen years. Results show a slight but not consistent age-to-age improvement, both for accuracy and time. However, because of the insufficient number of cases at each age level, no final conclusion can be drawn. (Boody, '26, p. 127)

Discussion: Boody thinks the test is not dependent upon age alone. Two of her youngest girls made no error. The girls showed a slightly better ability than did the boys in the finer discriminations of form, size and color. She found no racial differences. Dearborn found increased ability with age to differentiate geometrical form, but this in part he attributed to greater ease of movement. Boody's work revealed no contradictory findings. ('26 p. 127)

References:

Boody, '26

No. 141 Color Form Test

Devised by W. F. Dearborn.
Date: 1916
Handled by Psycho-Educational Clinic, Palfrey House, Harvard; and by Stoelting.
Price: $5.50 (Stoelting).
Material: Sixteen blocks, four squares, four triangles, four diamonds, four circles; one in each set being colored, red, blue, green and yellow.
Problem: To recognize and discriminate color from form. "Find all the blocks which are just the same shape as this one. Find all the blocks which are just the same color as this one."
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring method: Total time.
Standardization: This test is apparently not yet adaquately standardized, although Dearborn reports age scores for thirty-nine children. Range: from below four to beyond six years. (Dearborn *et al.*, '23, p. 39)
Discussion: "At these ages (four to nine) the child's sense of color is much better developed than is that of form." "The ability to discriminate simple forms becomes clear-cut shortly after the sixth year." (Dearborn, Anderson, Christiansen, '16, p. 447)
References:
Dearborn, Anderson, Christiansen, '16
Dearborn, Shaw and Lincoln, '23

No. 142 Comparison of Weights

Devised by A. Binet.
Date: 1908
Handled by Stoelting.
Price: $4.25
Material: As employed by Terman, these are small wooden blocks, similar in size and shape and color, but differing in weight. For Year V, they weigh respectively three and fifteen grams; for year IX, the weights are respectively three, six, nine, twelve, and fifteen grams. Or the material can be made up according to directions set down by Terman. ('16, p. 161)
Problem: At Year V, Subject is to tell which weight of two weights is the heavier; three trials. At Year IX, Subject is required to arrange the five weights in a graded descending order of heaviness; three trials.
Individual test; group adaptation difficult.
Directions: Verbal; pantomime adaptation possible.
Responses: Manipulative.
Time: A minute or two.
Scoring: Count as correct for the year level indicated, if successful in two of the three trials.
Standardization: Terman includes these tests in his Stanford-Binet scale. He reports a general agreement in assigning to Year V, the test which calls

for the comparison of two weights. The other test, that of placing the blocks in a decreasing series of weight, was originally placed by Binet in Year IX, and later in Year VIII. Other revisions have replaced it in the ninth year. The correct location will vary with the kind of weights used, the procedure and the method of scoring. Kuhlmann made the test easier by using weights 3, 9, 19, 27, 36, and 45 grams. By experimenting with two sets of boxes, Bobertag found that more errors occurred with the larger set. Terman points out also that requiring only one success in three trials would locate the test a year or two lower, while three successes would move it upward possibly two years. (Terman, '16, p. 237)

Discussion: Success on this test calls for a correct understanding of the task, and the ability to keep the goal in view; also selection of a suitable method of procedure. (Terman, '16, p. 237)

References:

Binet, '08 and '11
Wallin, '12
Yerkes, Bridges, Hardwick, '15
Whipple, '14
Porteus and Hill, '20
Kuhlman, '22
Yerkes and Foster, '23

No. 143 Copying of Geometrical Forms

Devised by A. Binet. Adapted by H. H. Goddard, R. M. Yerkes, H. A. Knox, L. M. Terman and F. Kuhlmann.

Date: 1908

Published by Houghton, Mifflin (with Stanford-Binet material); also separately by Stoelting; or can easily be made up.

Materials: Pencil, pen and ink; cards on which are drawn models of circle, square, diamond; each figure about one and a quarter inches wide.

Problem: To copy the model: at two and four years with pencil; at seven with pen.

Individual test; three trials; could be adapted to group testing.

Response: Non-verbal.

Time: A minute or two.

Scoring: One trial out of three for the circle and the square; for diamond, two trials out of three. Quality of drawing is measured by comparison with copies supplied on score cards with the Stanford-Binet material (for square and diamond).

Standardization: Kuhlmann placed the drawing of the circle in pencil at two years. Terman placed the drawing of the square in pencil at four years, and the drawing of the diamond with pen and ink at seven years in his Stanford-Binet. For detailed discussion as to variations in standardization, see Terman. ('16, p. 155 ff)

Discussion: Success in a test of this sort does not depend on drawing ability or previous training in drawing, but on the Subject's "appreciation of spatial relationships." The entire figure must be perceived and this visual impression must be transferred to paper with the aid of "a rather complex set of motor coordinations." It is these latter which seem to constitute the main difficulty, especially in the very early years. In commenting on how much a change in the angles, from square to diamond, has added to the difficulty of the test, Terman suggests that it would be worth while to devise and standardize still more difficult and complicated figures. (Terman, '16, p. 156f)

References:

Binet, '08 and '11
Goddard, '11
Knox, '14
Yerkes, Bridges and Hardwick, '15
Terman, '16
Kuhlmann, '22
Yerkes and Foster, '23

No. 144 Discrimination of Forms

Devised by F. Kuhlmann; modified by L. M. Terman.

Date: 1912

Handled by Houghton, Mifflin Co. (with Stanford-Binet material); and by Stoelting (separately).

Price: $.20 (Stoelting).

Material: A card on which ten geometrical forms are drawn; a similar set of such forms, cut out separately.

Problem: To match one of the "cut-out" forms with its mate on the card containing ten different forms.

Individual test; not readily adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Two or three minutes.

Scoring: The test is passed if seven out of ten choices are correct.

Standardization: This test was first standardized by Kuhlmann and placed by him at the four-year level. Terman's results substantiated this location. It is included in his Stanford-Binet scale.

Discussion: Failure on this test is more apt to be due to poor discriminative powers than to ability to follow directions. Like the formboard tests it calls for an ability "to compare and contrast successive visual perceptions of forms." Terman claims that the forms used in this test are meaningful to the young normal child of four, even though he has never heard of their names: triangles, squares, circles, and the like. Accurate perception of form and meaning at this age probably involves "a compound of such factors as appreciation of symmetry and direction, and discrimination of quantity and number." Certain demands on the child's powers of visual attentiveness are also made. He must not only attend; he must also compare one form with many others. Weakness here is indicative of a mental level of less than four years of age. (Terman, '16, p. 153)

References:

Kuhlmann, '12
Terman, '16
Porteus and Hill, '20

No. 145 Form Discrimination

Devised by C. B. Davenport.
Date: 1927
Published by the Eugenics Record Office.
Price: $1.00 per set, with key.
Material: Three booklets with pairs of geometric figures on each page, one of the pair being regular, the other distorted. Blank forms for recording response.
Problem: To decide whether each figure is "perfect" or "not perfect" geometrically.
Individual test; adaptable to group testing.
Directions: Verbal; pantomime possibly.
Responses: Verbal; non-verbal by adaptation.
Time: Ten seconds for each decision.
Scoring: Percentage correct.
Standardization: This test has not been standardized.
References:
Davenport, unpublished report

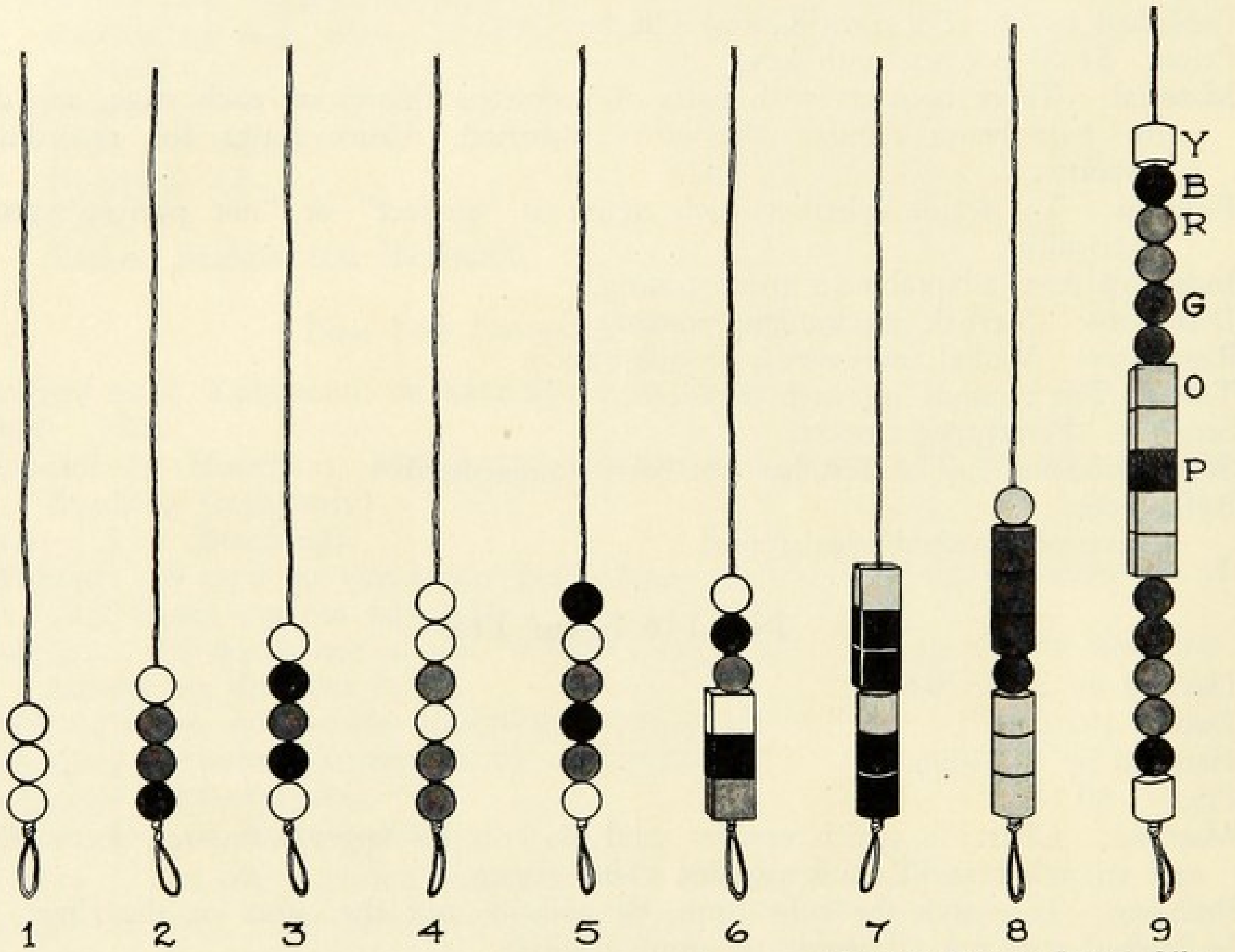
No. 146 Form Test

Devised by S. de Sanctis.
Date: 1910
Handled by Stoelting.
Price: $0.50
Material: Chart on which various solid black forms appear; squares, rectangles and triangles; small black wooden cube; screen.
Problem: To match the cube form, by pointing out the cubes on the chart.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Scoring: Time, mistakes, omissions.
Discussion: This test is part of an early scale proposed by de Sanctis for estimating degrees of mental defect. It has received very little attention.
References:
Whipple, '10 (pp. 469-473)
de Sanctis, '11

No. 147 Stringing Beads

Devised by John H. Long.
Date: 1929
Made up at the Eugenics Record Office.
Material: (1) Box of 1000 assorted beads in three forms: spherical, cylindrical and cubical; and six colors: purple, blue, green, yellow, orange, red; (2) cards, 1 x 3 centimeters, to attach at one end of string; (3) strings 25 centimeters long, threaded into large needle.

Problem: To reproduce from memory a model string of beads. Twelve models have been prepared, involving factors of color, number, sequence, shape differences, symmetry and pattern.



Courtesy Eugenics Record Office

Individual test; not adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: After first string has been copied without a time limit, five seconds are allowed to view strings 2 to 9 in succession. Eight seconds are allowed for viewing strings 10 to 12.

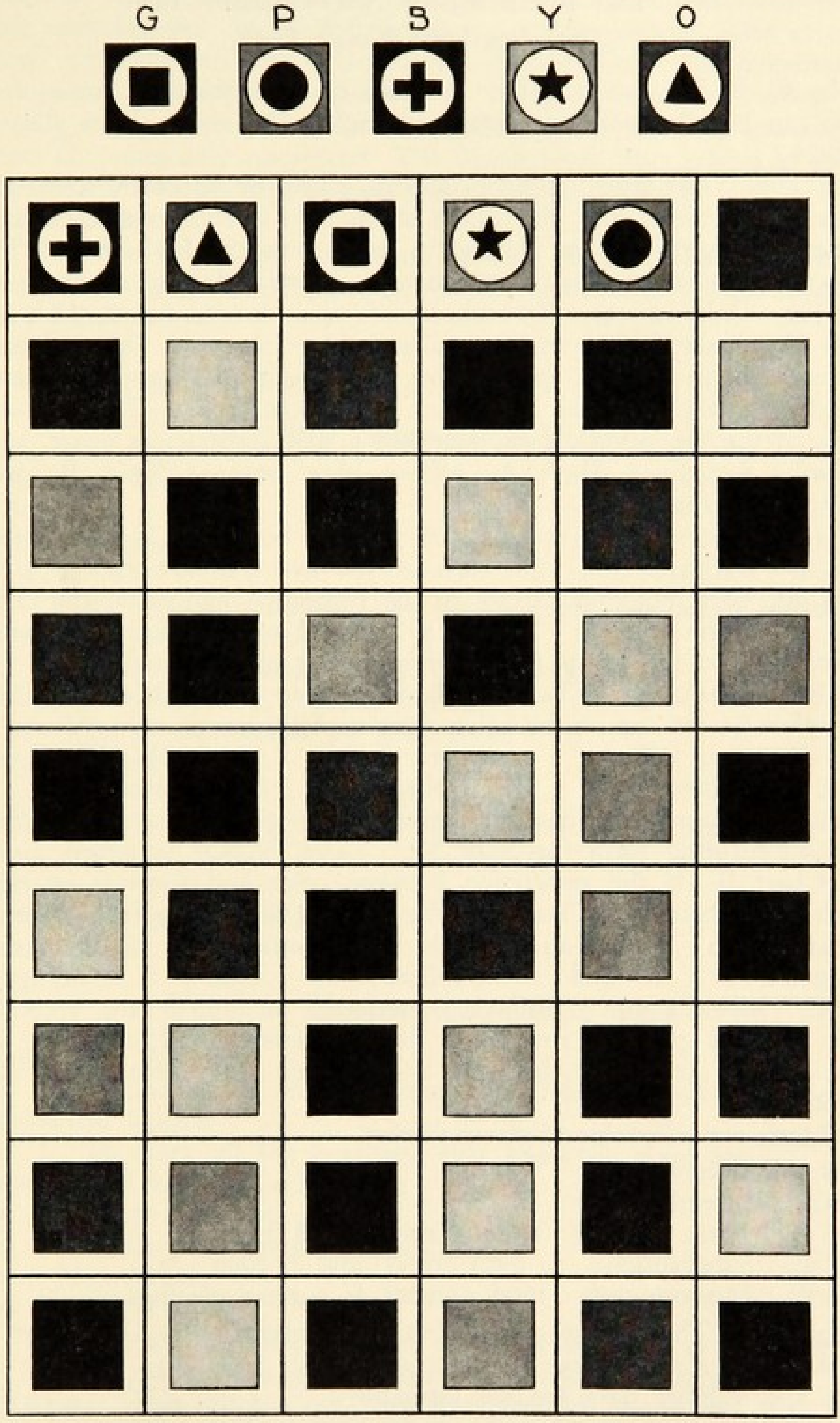
Scoring: Total number of strings correctly copied.

Standardization: This test has not been standardized.

Discussion: The qualities which seem to be brought out by this test are: (1) Ability to form and retain a unified visual mental image, and to analyze the elements of color, number, shape, sequence and pattern involved in the series. (2) Inventiveness in the use of cues, especially with the complex strings; (3) motor coordination; (4) perseveration; (5) auto-criticism; (6) ambition. (Davenport)

Reference:

Davenport, C. D., unpublished report



Courtesy Eugenics Record Office

No. 148 Form Color Substitution Test
(Including Form Sorting Test)

Devised by C. B. Davenport (after Woodworth and Wells). Some of this material can be obtained from Milton Bradley Co., New York City.

Date: 1929

Material: (1) Unpainted, circular wooden chips, or checkers, seven-eighths of an inch in diameter and one-fourth of an inch in thickness; each having one geometrical figure drawn solidly in India ink on both sides. Five such geometric figures are used: star, cross, circle, triangle and square.

(2) Color sheets. Two sheets of white bristol board, 12 x 20 inches. Sheet A contains three key units and fifty-four test units; sheet B con- gummed circles of red, green, blue, yellow and orange. The circles are placed in nine rows of six units each in random order. The key units form a row at the top of the sheet, one and a half centimeters apart, and in the following sequence: Sheet A: yellow, blue, orange; Sheet B: green, red, yellow, blue, orange.

(3) Box for form-sorting: A wooden box, 40 x 15 x 3½ centimeters, with five compartments. On each of the five compartments is a figure in India ink, corresponding exactly to the respective figures on the chips.

Problem: A. Form sorting: Subject is required to sort the chips, scattered at random before him, into their appropriate compartments.

B. Substitution: Chart A. Experimenter places checkers on the three key colors at the top of the color sheet as follows:

cross, on yellow	square, on blue	circle, on orange

After a demonstration has been given, Subject is required to "finish" the page, substituting form for color.

Chart B. If the association between color and form is established for Chart A, proceed with Chart B The position of the key forms is as follows:

triangle, on green	star, on red	cross, on yellow	square, on blue	circle, on orange

C. Memory test: Subject is required to fill in the second, rather than the first practice row, after having removed all checkers, including those of the key row; this test to be given after completion of Charts A and B.

Individual test; not adaptable to group testing.

Directions: Pantomime.

Responses: Non-verbal.

Scoring: A. Form sorting: 1, total errors; and 2, time.
B. Substitution: 1, total errors; and 2, time.
C. Memory test: total number of correct placings.

Standardization: This test was tried out experimentally by Davenport, but has not yet been standardized.

Reliability and validity data are not available.

Discussion: Davenport considers the Form Sorting Test comparable with Pyle's Card Sorting and Marble Sorting Test. The test calls for such psychological factors as attention, speed of perception and form discrimination, visio-motor co-ordination, speed of developing a motor position habit in associating symbol with compartment. Davenport considers that the Substitution Test "tests one of the essential elements of language, *viz,* the ability to apply correctly a set of systematized and arbitrary symbols; to associate sets of objects not essentially connected. The Substitution Test relates to an extremely simple type of association, one independent of culture and open to direct interpretation."

"The Memory Test," he states, "gives a valuable check on previous observations to determine whether or not the association has been established." (unpublished report).

References:

Davenport, unpublished report

No. 149 Geometrical Forms

Devised by S. de Sanctis.
Date: 1910
Handled by Stoelting.
Price: $0.75
Material: Three square pyramids, two inches high; five small wooden cubes; two wooden parallelopipeds, 2½ x ½ inches.
Problem: To pick out the cubes from the collection.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Scoring: Time.
Discussion: This test is part of a scale proposed by de Sanctis for detecting various grades of mental defect. (de Sanctis, '11)
References:
Whipple, '10 (pp. 469-473)
de Sanctis, '11

No. 150 Geometrical Forms Test

Devised by M. T. Whitley.
Material: Pencil and paper on which are drawn sixteen rows of different kinds of geometric figures.
Problem: 1. To mark all the hexagons that have the point up. 2. To mark all semi-circles with flat side up.
Individual or group test; two parts.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Time: A minute or two.
Scoring: Number of seconds plus three for each omission in first test.
Number of seconds plus six for each omission in second test.
Standardization: This test has not been standardized, but it was used by Simpson on his adult group of thirty-seven men: seventeen superior and twenty inferior in intelligence.

Reliability: r = .90, first and second trials, on good and poor groups. (Simpson, '12, p. 43)
Validity: r = .40 with the other eleven tests in Simpson's battery. (Simpson, '12, p. 66)
References:
Whitley, '11 (p. 69-70)
Simpson, '12

No. 151 Identification of Forms Test

Devised by H. T. Woolley.
Date: 1920
Handled by Stoelting.
Price: $17.50
Material: Two sets of similar forms, ten to a set. They are made of varnished quarter-inch board; one set is mounted on a blackboard, 10½ x 21 x ⅜ inches; a small table screen.
Problem: After being allowed to feel one of the shapes behind the screen, the Subject is required to identify the shape upon removal of the screen.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Limit of ten seconds for each "feeling" period.
Scoring: 12½ percent for each correct choice.
Standardization: This test was standardized by Dewey, Child and Ruml on 500 Jewish cases. Norms are expressed in terms of percents and are distributed for sex and age. S. D.'s are given. Age range: from nine to thirteen. The age norms, however, are not discriminative. ('20, p. 30-31 and 79)
Discussion: Dewey, Child and Ruml state that Woolley has here attempted to devise a test which would involve "the transfer from a kinaesthetic to a visual impression," calling for a "type of discrimination and comparison which is common in life." ('20, p. 31) Apparently this test has not enlisted the interests of other investigators, for as Bronner, Healy, *et al*, state: "So little has been done with this test that it is impossible to know whether the originators' aims have been realized". ('27, p. 224)
References:
Dewey, Child, Ruml, '20
Bronner, Healy, Lowe and Shimberg, '27

No. 152 Modification Test

Devised by A. R. Abelson; modified by H. A. Ruger.
Date: 1917
Materials: A series of outline drawings containing circles, squares and triangles, interlocking one another in various ways, some simple, some complicated. Abelson offered geometrical designs, but the Ruger modification substituted familiar animals for these forms, as many immigrants, literate and illiterate, were unfamiliar with these geometric forms. Combinations were

as follows: 1, pig and chicken; 2, pig and chicken and man; 3, two pigs and chicken and man; 4, two pigs, two chickens and man, and so on.

Problem: To point to a spot on a certain picture combination that will satisfy directions.

Directions: Verbal; pantomime doubtful.

Responses: Non-verbal.

Time: Two or three minutes.

Scoring: Number correct.

Standardization: This test is not adequately standardized, but was reported upon by Mullan, who used it on 295 cases of literate and illiterate aliens, children and adults. He states that seventy-five per cent of every group were successful with the designs containing at least four animals. No difference existed in this test as regards age, the ten- and eleven-year-old children doing as well as adults. (Mullan, '17, p. 109)

Discussion: Because this test presented a new and complex situation to the majority of immigrants, Mullan classified it among the problem or reasoning tests. Literates and illiterates did about equally well on it. "In view of these facts" adds Mullan, "it might be well to claim that this test is one of the best tests of the entire investigation for bringing forth native ability." (Mullan, '17, p. 109)

Reference:

Mullan, '17

No. 153 Visual Comparison Test

Devised by H. A. Knox.

Date: 1913

Published by Stoelting.

Price: $0.25

Material: A card containing five groups of figures.

Problem: To designate or "pair" the five pairs in twenty-five seconds.

1. Leaf section, as above.
2. Envelope section, with five pairs; different design. Eighteen seconds.
3. Face line section—to point out "sad" faces. Twenty seconds, three trials.
4. Moon section: To point out four moons looking to left. Fourteen seconds.
5. Key section: To find out the "nearest like" the picture shown. Time not worked out.

Individual test; not adaptable to group testing.

Directions: Verbal; doubtful if all could be adapted to pantomime.

Responses: Non-verbal.

Time: See above, under Problem.

Scoring: Not given.

Standardization: Knox reported this test, but did not pretend to standardize it He used it for illiterates over twelve years of age.

References:

Knox, '13 and '14

No. 154 Recognizing Forms Test

Devised by M. T. Whitley.
Date: 1912
Handled by Stoelting.
Price: $3.20 per hundred sets.
Material: Two papers on which twenty geometric forms are drawn and exposed one, and one and a quarter minutes for learning; two recognition blanks, each containing forty forms, some of which are identical with the forms in the key sheets.
Problem: To identify and mark all the forms seen on the key sheets.
Individual or group test; two parts.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal.
Time: One minute in which to study; no time limit for marking.
Scoring: One point for each form correctly marked.
One point deducted for each form incorrectly marked. (Simpson, '12, p. 14)
Standardization: This test has not been standardized, but was used by Simpson ('12) on his group of thirty-seven adults, consisting of seventeen superior and twenty inferior men.
Reliability: r = .40, first and second trials, on Simpson's thirty-seven cases. (Simpson, '12, p. 45)
Validity: r = .41 with the other eleven tests in Simpson's battery. (Simpson, '12, p. 66)
Discussion: Simpson says: "The variability of the two trials is so great as to swamp small differences, such as occur within either group and to show only very crudely, the larger differences between an individual in the good, and one in the poor group." ('12, p. 45)
References:
Whitley, '11, (p. 53)
Simpson, '12
Poffenberger, '16 (Ex. 11)

No. 155 Weight Illusion Blocks

Devised by De Moor.
Date: 1913
Handled by Stoelting.
Price: De Moor's $7.00; Mayer's $7.50; Gilbert's $18.00; Seashore's $68.00; Whipple's $14.00; Kline's $8.25; (at Marietta $7.50.)
Material: Two blocks of poplar wood, 4.5 x 11.0 x 7.5 centimeters and 4.5 x 11. going up by 5 gram increments. For other descriptions see also Seashore ('99) and Scripture ('96).
Problem: To tell which of two blocks is the heavier.
Individual test; not adaptable to group testing.
Directions: Verbal; pantomime doubtful.

Responses: Manipulative and verbal.

Standardization: Norms are expressed in terms of age by Doll who reported giving the test to 345 feeble-minded children. Range: four to eight years. (Doll '13) The relation of intelligence to suggestibility has not been studied as carefully as the problem warrants.. (Whipple, '15, p. 596)

Discussion: Doll considers that this test of "illusion of size-weight is a test of mental development." (Doll, '13, p. 149)

References:

Doll, '13
Whipple, '15 (pp. 590-598)
Kline and Kline, '27 (Test 80)

No. 156 Diagram Drawing

Devised by Binet-Simon; modified by F. W. Steacy.

Date: 1919

Material: Two sheets of sixty diagrams. The first sheet contains thirty figures. Each of these is the upper left-hand quarter of the diagram to be drawn, and bears its distinctive number. The second sheet has forty dotted outlines all alike, on which the pupils can draw the required diagrams. The diagrams are 1.5 inches long and the same in width.

Problem: To draw the diagram in forms provided and to write in the numbers according to the pattern.

Individual or group test; two forms.

Directions: Verbal; pantomime adaptation doubtful.

Responses: Non-verbal.

Time: Thirty minutes for each series.

Scoring: One mark is allowed for each diagram properly drawn. The highest possible score for each series is thirty marks.

Standardization: This test is not standardized, but was worked with experimentally by Steacy on thirty-one boys and thirty-eight girls in the sixth grade of a New York public school. Average deviations are given for the group, in composite.

Validity: r = .34.

Reliability: r = .80 with the other tests in the Steacy battery, exclusive of handwriting. "The large correlation is due probably to large average deviations and due also to the fact that both chance and practice have but little effect on the subject's score." (Steacy, '19, p. 40)

Discussion: Steacy thinks that this test "requires synthetic ability. The subject must be able to understand and to follow a pattern and to carry a spatial scheme in mind. A knowledge of words beyond the instructions is unnecessary." (Steacy, '19, p. 40).

References:

de Fursac, '11, p. 437-8
Goddard, '11
Steacy, '19

No. 157 Diagram Matching

Devised by F. W. Steacy, after A. Binet and T. Simon.
Date: 1919
Material: Two sheets of paper, each sheet containing twenty diagrams. The diagrams on the second sheet are duplicates of the upper left hand quarters of the diagrams on the first sheet. These quarters are numbered consecutively, but scattered in a different order from their originating models on the first sheet. These models are symmetrical as far as their quarters are concerned.
Problem: To identify the diagrams, selecting the whole diagram corresponding to the upper left hand quarter, and to write in the whole diagram its corresponding number.
Individual or group test; two forms.
Directions: Verbal; doubtful if they could be adapted to pantomime.
Responses: Non-verbal.
Time: Fourteen minutes for each series.
Scoring: One mark is allowed for each diagram, properly identified. The highest possible score is twenty marks for each series.
Standardization: This test was worked with experimentally by Steacy on a group of thirty-one boys and thirty-eight girls, in the sixth grade of a New York public school. He has distributed his scores by sex, giving average deviations.
Reliability: r=.52
Validity: r=.48 with the other tests in the Steacy battery, exclusive of handwriting. (Steacy '19, p. 40)
Discussion: Steacy thinks that the test calls for analytic ability; also ability to recognize identities, and to hold spatial schemes in mind. The Subject will separate off the essential part of the whole diagram, matching it with the given quarter. He needs to know only enough words to follow the instructions and enough numbers to enable him to write the identification numbers. (Steacy '19, p. 40)
References:

de Fursac, '11 p. 437-8
Goddard, '11 p. 9
Steacy, '19

No. 158 Patience or Divided Triangles

(See also Cornell's Constructive Problems Test. No. 163)

Devised by A. Binet.
Date: 1911
Handled by Stoelting.
Price: $0.10 apiece; or can easily be made up.
Material: Two rectangular cards, each 2 x 3 inches, one of which is divided into two triangles by cutting along its diagonal.
Problem: To put together the two triangles so that they will look like the uncut card.
Individual or group test.
Directions: Verbal; pantomime adaptation possible.

Responses: Manipulative.

Time: A minute or two.

Scoring: There must be two successes out of three trials.

Standardization: Terman says of this test that it is "rather easy for year V, though plainly somewhat too difficult for year IV". He places it at year V in the Stanford-Revision of the Binet-Simon scale. ('16, p. 171)

Discussion: Binet pointed out that the Subject must keep in mind the end to be attained, must not lose this directing idea, and must be able to make a comparative judgment when he has finished his task to decide whether his combination compares favorably with the model. He called it a "test of patience" because of the need to persist if the end is to be achieved. Terman thinks it would be well worth while to embody this principle in a new and more difficult test, but warns against including the puzzling features of the usual games of anagrams. He suggests that this test has something in common with the form-board test, the Ebbinghaus test, and the dissected sentences test. (Terman '16, p. 171)

References:

Binet, '08, '11
Goddard, '11
Terman, '16

XII
SPATIAL RELATIONS

No. 159 Geometrical Construction Test

Devised by R. M. Yerkes and J. C. Foster.
Date: 1923
Handled by Stoelting.
Price: 10 cents.
Material: Two heavy cardboards, on one of which a square is drawn, and on the other, six parts which, fitted together would make up a similar square.
Problem: To show with a pencil how the six parts can be fitted together to just fill the square.
Individual test; two trials allowed; adaptable to group testing.
Directions: Verbal; pantomime with difficulty.
Responses: Non-verbal.
Time: First trial, two minutes. Second trial, two minutes.
Scoring: Points according to time and trial needed. (Yerkes and Foster '23, p. 113)
Standardization: This test was standardized by Yerkes and Foster on fifteen grammar school children, twenty-five working children and eighty college students and professional people.
Discussion: This test is a part of the Yerkes-Foster revision of the Binet-Simon scale, and falls in the "pre-adolescent group" of tests. (Yerkes and Foster '23, p. 116f)
References: Yerkes and Foster, '23

No. 160 Paper Form Board

Based on Army Beta, (Geometrical Construction Test). Revised by D. G. Paterson, *et al.*
Date: 1920; revised 1928
Handled by Marietta; also by Stoelting.
Price: $3.00 per hundred (Marietta); $3.35 per hundred (Stoelting).
Material: Series A and B. Two four-page blanks on which are printed fifty-six items, each consisting of one large figure and two or more smaller ones, which are segments of the large one. The items are arranged in order of difficulty.
Problem: To indicate by drawing lines in the large figure how the small ones could be fitted into it.
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Manipulative.
Time: Fifteen minutes for each series.
Scoring: Number of figures correctly marked.
Standardization: This test has been very carefully standardized by workers at the University of Minnesota. Norms are provided for each sex separately,

and together; they are distributed by age and grade and in percentiles. Range: ages eleven to twenty-one; grade VII through university groups. (Manual, in mimeograph, p. 40 ff)

Reliability: r = .90 (Anderson '28b)

Vilidity: Approximately the same as for the Minnesota Assembly Test. (See Test No. 68 (Anderson '28b).

Discussion: See discussion on Minnesota Assembly Tests. (Nos. 68 and 171).

References:

Rogers, '24

L. D. Anderson, '26, '27, '28a and '28b

Edgerton and Paterson, '26

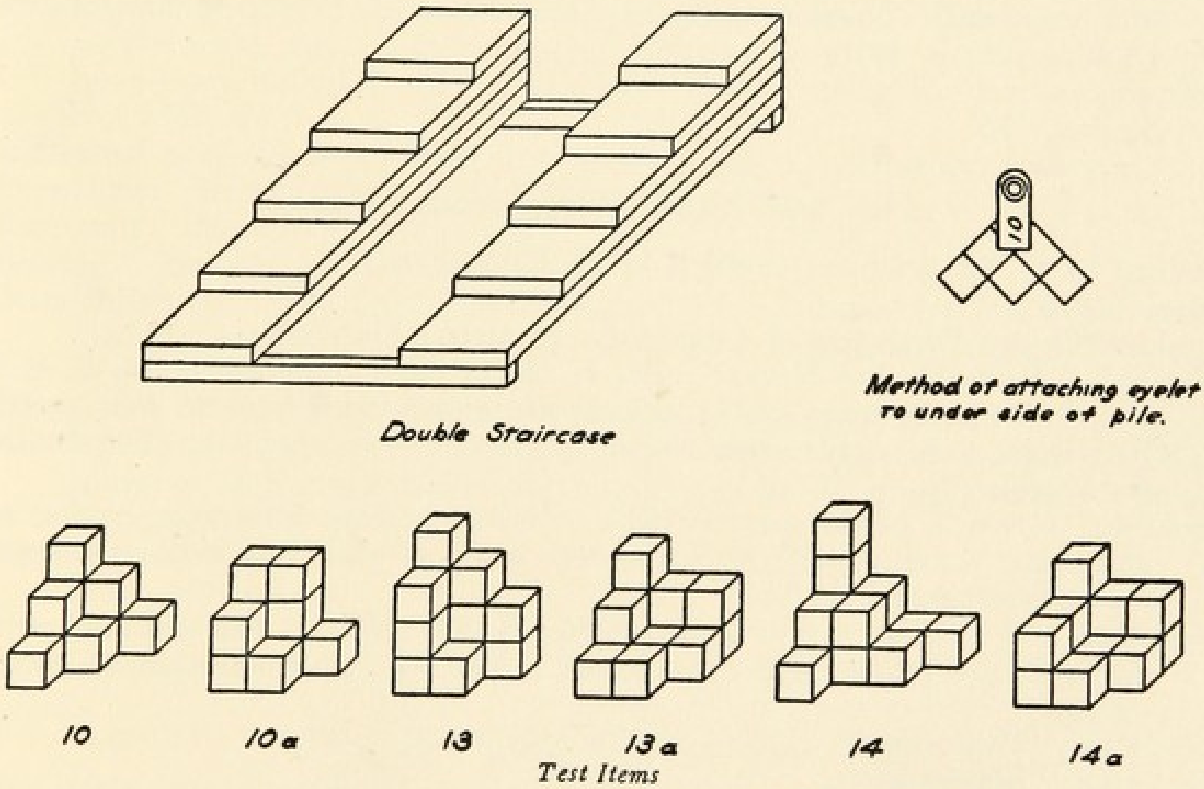
Paterson, '28

Hubbard, '28

Carter, '28

Paterson, Elliott, Anderson, Toops and Heidbreder. To be published by University of Minnesota Press.

No. 161 Arrangement of Cube Piles



Reproduced by courtesy of Princeton Univ. Press

Devised by P. C. Squires.

Date: 1926

Produced by the Psychological Laboratory, Princeton University.

Materials: Thirteen cube piles used for demonstration purposes and seventeen piles in the test series; two boards, 16 x 6½ inches; a double "staircase" 5 x 10 inches. For details, see Squires ('26, p. 44ff)

Problem: Given a certain number of cube piles, to arrange the piles by "twins" in order of size upon opposite steps of the "staircase".
Individual test: Three practice piles, two test boards, each containing two rows for cubes. The test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Time for each of the two boards; moves; successes.
Standardization: Squires tried out this test on fifty Princeton seniors and fifty feeble-minded males, but norms have not been established. The test appears useful to test high grade intelligence.
Validity: r = .11 (successes) and —.32 (time) for the Princeton group, against a criterion of the higher score on either the Thorndike or Princeton intelligence tests.

r = .25 (successes) and .27 (time) for the feeble-minded group; criterion was Stanford-Binet mental age. (Squires '26, p. 184)

Discussion: Squires reports that this test did not turn out as well as did the Single Cube Piles Test (No. 177). He advises that the number of cubes for either test be limited to twenty, and that a third board should be added to this one, which should contain very easy items for the low-grade subjects. (Squires '26, p. 191f)
References:
Squires, '26

No. 162 Block Construction Test

Devised by C. C. Brigham; modified by P. C. Squires.
Date: 1926
Produced in the Psychological Laboratory, Princeton University.
Materials: Series A, five blocks: Series B, blocks whose numbers are 1, 2, 4, 5 and 6; Series C, seven blocks. Blocks are glued together or in solid wood. Forty-eight loose construction blocks, boards, and two abacuses. (For details, see Squires '26, p. 77 ff.)
Problem: Given a block, together with a board bearing four type blocks, to determine, being allowed N blocks, how many of each type-block enter into the construction of the block.
Individual test: Series A, five items. Series B, four practice sets and three sets. Series C, two practice sets and five sets. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Time on entire block; moves; successes.
Standardization: Squires experimented with this test on fifty Princeton seniors, and fifty feeble-minded males at Vineland. Norms are not established.
Validity: r = .33 (successes) and —.46 (time); for the Princeton group, against a criterion of the higher score on either the Thorndike or Princeton intelligence examination.

r = .35 (successes) and .27 (time) for the feeble-minded group: criterion, the Stanford-Binet mental age.

Discussion: This test covers a wide range of mental capacity from very simple to highly developed mental processes. "The real object of this test is to determine the capacity of an individual to analyze and to construct 'mentally' without having recourse to actual construction by means of loose blocks . . . . The task in the case of this test involves essentially the type of factor which may be termed qualitative." (Squires '26, p. 89) This test ranked high in the composite of eight tests for both the Princeton and Vineland groups. (p. 192)

References:

Squires, '26

No. 163 Constructive Problems Test

Devised, in part, by A. Binet; modified by C. B. Cornell.

Date: 1917

Material: Pieces of paper and cardboard.

Problem: Problems 1, 2 and 3 are cardboard puzzles: fitting triangles together to produce a form which is placed before the Subject. Problem 4 is Binet's paper-cutting test; 5, is Binet's interchanged hands of the clock test; 6, is Binet's reversed triangles; 7, is Binet's paper cutting for adults. (See our No. 158)

Individual test; seven problems; not readily adaptable to group testing.

Directions: Verbal; some are adaptable to pantomime.

Responses: Non-verbal and verbal.

Scoring: Number of correct solutions.

Standardization: These tests were standardized by Cornell on 550 public school children, ranging from six to fourteen years. Norms are expressed in terms of age.

Discussion: "This is a group of seven problems designed to test the judgment, ingenuity and concentration of the child." All of the tests are not necessarily given to each child, unless he shows promise of passing them. (Cornell '17, p. 546)

References:

Binet, '08 and '11
Terman, '16
Cornell, '17

No. 164 Cube Construction Test

Devised by E. A. Doll; modified by the Army psychologists.

Date: 1918

Handled by Stoelting.

Material: Three sets of cubes and models as follows:

1. Block, 3 x 3 x 1 inches, marked into one-inch cubes, painted on four sides, but not on top or bottom. Nine one-inch cubes, one unpainted, four painted on one side, and four painted on two sides.

2. Block, 3 x 3 x 1 inches, but painted on top; nine one-inch cubes, as for Part I.
3. Block, 2 x 2 x 2 inches, marked into one-inch cubes and unpainted; eight one-inch cubes painted on three sides.

Problem: To put the blocks together to form a large cube, "painted all over on the outside".

Individual test. Three different problems; to build up three models partially painted. This test is not adaptable in its concrete form to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: Six minutes.

Scoring: Sum of points for moves and time according to table of credits. (Yoakum and Yerkes '20, p. 107)

Standardization: This test is not adequately standardized.

Validity:

r = .633 with Stanford-Binet mental age on 260 army men.

r = .819 with Army Performance Scale on 260 army men. (Yerkes, (Ed) '21, p. 404)

r = .75 with foreman's ratings on workers' ability to learn their prescribed work. (Link '19, p. 124 f)

Discussion: To test practical ability "Cube construction tests are particularly discriminative in this regard because success depends not on mere imitation, but on the ability to grasp the spatial relations obtaining between the cubes." (Bronner, Healy, Lowe and Shimberg '27, p. 212)

See also the discussion on Painted Cube Test No. 174

References:

Doll, '17
Link, '19
Yoakum and Yerkes, '20 (p. 105ff)
Yerkes, (Ed.) '21
Gaw, '25a and '25b
Dodd, '26a, and '26b
Squires, '26
Bronner, Healy, Lowe and Shimberg, '27

No. 165 Cube Test

Devised by E. A. Doll.

Date: 1926

Material: A two-inch cube painted red all over on the outside and eight one-inch cubes painted on three sides that form a corner.

Problem: To build up a two-inch cube that will be painted red on all its sides.

Individual or group test.

Directions: Verbal or pantomime.

Responses: Manipulative.

Scoring: Time.

Standardization: This test is not adequately standardized, but a provisional standardization is offered by Doll, based on the results of giving it to 185 inmates of a state reformatory, all being over sixteen years of age. The norms are expressed in terms of time, standardized against the Stanford-Binet mental ages, and the range is from eight to sixteen years. Point scores are also offered, and percentiles for grouped ages: eight-nine, ten-twelve, thirteen-sixteen.

Validity: r = .53 against criterion of Eight-Piece Picture test. No. 82.
r = .55 against criterion of Twelve-Piece Picture test. No. 83.
r = .44 against criterion of Stanford-Binet test.

Discussion: This test is a modification of the Three-inch Painted Cube. Doll considers it of definite experimental value in testing illiterates and foreigners. It lends itself well to group testing, as the factor of copying does not count in a test of this sort. (Doll '26)

References:
Yerkes, '21 p. 365
Doll, '26

No. 166 Drawing Lines

Devised by E. L. Thorndike (?)

Date: 1912

Material: Three lines drawn on paper, 100, 75 and 50 millimeters respectively.

Problem: To draw three lines to the right of the model, each equal to the model. Each trial is covered up before next is attempted.

Individual or group test.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: A minute or two.

Scoring: Sum of the deviations, plus and minus, from the standard length. (p. 15)

Standardization: This test is not standardized, but was worked with experimentally by Simpson on a group of thirty-seven cases, seventeen of them being superior adults, and twenty being inferior.

Reliability: r = .72 between first four trials and last four trials. Simpson says that it would have been better to have taken four lengths instead of three. (p. 46)

Validity: r = .13 with the other eleven tests in the Simpson battery. (Simpson '12, p. 66)

Discussion: "It was evident to the experimenter that it was in part a test of painstaking and patience". (Simpson '12, p. 46)

Reference:
Simpson, '12

No. 167 Estimating Lengths

Devised by A. Binet.
Date: 1908
Handled by Stoelting.
Price: $0.60
Material: Different experimenters have used different sets of materials. Goddard's

1908 set consisted of thirty-three cards with lines of various lengths, but later, in 1911, he used six cards, three of which contained lines of different lengths and three lines of the same length. Yerkes ('15) has a similar set. Terman's ('16) modification consisted of two lines. Simpson ('12) used four sets of two pairs of lines, drawn on cardboard, 100 and 108 mm; 100 and 106 mm; 100 and 104 mm; and 100 and 102 mm.

Problem: To estimate which of each pair of lines is the longer. (Simpson '12)

Individual test; could be adapted to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Verbal; non-verbal by adaptation.

Time: A minute or two.

Scoring: Number of correct judgments. Simpson points out a possible difference in effect which relative position in a short series, as contrasted with a long series, would have on scoring. (p. 46)

Standardization: Each standardization varies, of course, with the materials used. Simpson tried out his test experimentally on thirty-seven adults, seventeen superior and twenty low grade men, but he did not standardize it. Goddard placed his adaptation in the twelve-year-old group. Terman places comparison of two lines at the four-year level.

Reliability: r = .48, first half against second half for 37 cases. (Simpson '12, p. 46)

Validity: r = .26, against the eleven other tests in Simpson's battery. (p. 66) Simpson suggests that it would be better to have made the differences 1, 2, 3, 4 mm. respectively instead of 2, 4, 6, 8 mm. (Simpson '12, p. 46)

Discussion: "The very considerable differences between the correlation of random halves of the same test with each other, and the correlation of the two tests (i. e. Drawing Lines and Estimating Lengths) with each other, make it clear that there are very appreciably different factors involved in the two tests as given." (Simpson '12, p. 66)

Terman points out that Binet has noted that success in this test depends on the comprehension of the verbal directions rather than on actual discrimination of length. However, since three trials are required in the Stanford Revision, willingness to attend becomes an important factor. Terman adds that it is "exactly in such voluntary control of mental processes that we find one of the most characteristic differences between bright and dull, or mature and immature subjects." ('16, p. 151)

References:

Goddard, '08 and '11
Whipple, '10
Simpson, '12
Town, '13
Whipple, '15
Yerkes, Bridges, Hardwick, '15
Terman, '16
Porteus and Hill, '20
Kuhlmann, '22
Yerkes and Foster, '23

No. 168 Hands Test

Devised by L. L. Thurstone.
Date: 1925
Published by Stoelting.
Price: \$2.50 per hundred.
Material: Forty-nine drawings of right and left hands in different positions.
Problem: To indicate whether it is the right or left hand which is represented in each drawing.
Individual or group test.
Directions: Verbal; pantomime by adaptation.
Responses: Non-verbal, consisting of check marks to indicate whether each picture represents a right or a left hand.
Time: Three minutes.
Scoring: Number correct.
Standardization: This test was standardized by Thurstone on 238 college male students, and 114 college female students. Scores are distributed by percentiles. The test was used as part of an experimental study of three-dimensional space imagery. It is not intended to be used by itself, but in conjunction with other tests of space thinking, it may serve to differentiate between verbal and visual space thinking. It has not been validated for use by itself. (Thurstone, letter to writers)
Discussion: Bronner, Healy, *et al,* speaking of this and other tests, such as Spatial Relations, say: "They all seem to involve visual mental representation, though much work in standardization, intercorrelations with objective criteria is necessary to determine what factors are subsumed under the blanket term, *visual imagery.*" ('27, p. 231)
References:
Payne, '25
Squires, '26
Bronner, Healy, Lowe and Shimberg, '27 (illus. p. 38)
Thurstone, letter to writers

No. 169 Inverted Animals

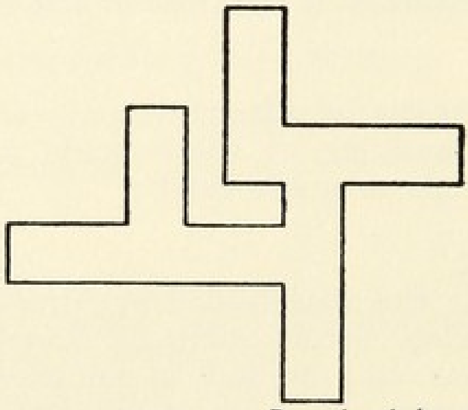
Devised by U. S. Public Health Service.
Date: 1917
Material: Pictures of five animals: horse, cow, donkey, pig, dog, presented upside down "in order to see if this baffling would interfere with immigrant's perceptive powers".
Problem: To name the animals.
Individual or group test.
Directions: Verbal; pantomime difficult.
Responses: Verbal; non-verbal by adaptation.
Time: A minute or two.
Scoring: A mistake in naming one animal constitutes total failure on the test.
Standardization: Mullan reported the results of giving this test to forty literate and illiterate aliens, but he made no claim to standardization. He states

that successes and failures ranged all the way from three-year-olds to adult aliens, literate and illiterate.

Reference:

Mullan, '17

No. 170 Inverted Figures Test



Reproduced by courtesy of Little, Brown & Co.

Devised by F. W. Ellis; modified by M. E. Goudge and Harry W. Crane.
Date: 1927
Published by Stoelting.
Price: $0.30
Material: A set of three cards containing geometrical drawings.
Problem. With a card before him, Subject must draw the figure as it would look if inverted.
Individual test: Three problems; could be adapted to group testing.
Directions: Specific; if pantomime, allow a sample form for preliminary practice.
Responses: Non-verbal.
Standardization: This test, which is suitable for older subjects, is not yet standardized.
Reliability and validity data now being prepared by Goudge.
References:

Kuhlmann, '22
Bronner, Healy, Lowe and Shimberg, '27
Proceedings, F. L. Wells, '30

No. 171 Minnesota Spatial Relations Test

Devised by H. C. Link; revised by D. A. Paterson, *et al.*
Date: 1919; revised 1928
Handled by Marietta; also by Stoelting.
Price: Boards A and B, $17.00; C and D, $17.00. Both $34.00 (Marietta); $35.00 (Stoelting)
Material: Two pairs of boards, each containing fifty-four cut-outs. The boards have no back base, so when the board is lifted from the table, the blocks remain on the table. In Link's test, the board was inverted, thus placing

blocks upside down. There is one set of blocks for boards A and B, and another set for Boards C and D.

Problem: With one hand only to place into the second board the blocks which have been left on the table after the Examiner has removed the first board. Four problems.

Individual test; not adaptable to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Time: A minute or two.

Scoring: Time in seconds, and errors.

Standardization: This test has been very carefully standardized by workers in the University of Minnesota. Percentile norms are provided separately for each sex, and for each age; also for time scores; also for error scores; also for each pair of boards; also for long form and short form of the test. Range: age from eleven through twenty-one years and above; grade, from VII through university engineering students. (Manual in mimeograph)

Reliability: Lower than for the Minnesota Assembly and Paper Formboard tests. Nos. 68 and 160. (L. D. Anderson '28b)

Validity: Approximately the same as for the Minnesota Assembly and Paper Formboard tests. (Anderson '28b)

Discussion: See the Minnesota Assembly Test No. 68.

References:

Link, '19

H. W. Rogers, '24

L. D. Anderson, '26, '27, '28a and '28b

Edgerton and Paterson, '26

Paterson, '28

Hubbard, '28

Carter, '28

Paterson, Elliott, Anderson, Toops and Heidbreder, to be published by Univ. of Minn. Press.

No. 172 Paper-Cutting Test

(See also Cornell's Constructive Problems Test. No. 163)

Devised by A. Binet.

Date: 1908

Material: Two small pieces of paper, similar in size and shape; pair of scissors.

Problem: Examiner folds paper two ways, and then cuts out notch on folded edge. Subject is then required to draw, from imagination, what the other paper would look like if treated in the same way.

Individual test; could be adapted to group testing.

Directions: Verbal; pantomime by adaptation.

Responses: Manipulative.

Scoring: Correct pencil reproduction of creases and holes.

Time: Two or three minutes.

Standardization: Binet first placed this test in Year XIII, but later shifted it to the adult level. Goddard and Terman have retained it there, but Kuhlmann places it in year XV. Variations in procedure and scoring affect its placement. As given in the Stanford-Binet, it is passed by about one-third of "average adults" and by a large majority of "superior adults". (Terman '16, p. 339)

Discussion: As indicated in the foregoing paragraph, this test forms one of the elements in the Stanford revision of the Binet-Simon scale. Terman thinks that it demands constructive visual imagination. He noted that his unschooled Subjects often did better than his high school or college students of the same mental level, and finds that solutions are seldom reached as a result of "logical mathematical thinking". The Subject must be able to picture to himself what the paper would look like when unfolded, both in regard to creases and to number and location of holes. (Terman '16, p. 339)

References:

Binet, '08 and '11
Goddard, '11
Terman, '16
Cornell, '17
Kuhlmann, '22

No. 173 Painted Cube Construction

Devised by P. C. Squires; after Goddard.

Date: 1926

Produced by the Psychological Laboratory, Princeton University.

Materials: Nine test cubes and one practice cube. The practice cube is composed of eight one-inch cubes; no surface is red. The other cubes have varying numbers of sides painted red. Twenty-four one-inch cubes, also varying as to painted sides. For details, see Squires. ('26, p. 54 ff)

Problem: Given a cube composed of smaller cubes and which has a certain number of its faces colored red, to determine how many of the component cubes are red on three, two, one and no sides.

Individual test: One practice item; three test sets, with nine items. This test is not adaptable to group testing.

Directions: Pantomime.

Responses: Manipulative.

Time: No time limit.

Scoring: Time; moves, three kinds; successes.

Standardization: Norms are not established, but Squires reports this test suitable for high grade adults. He established it on fifty Princeton seniors and fifty feeble-minded males.

Validity:

r = .14 (for successes) and —.36 (time) with Princeton group, criterion being the higher score on Thorndike or Princeton intelligence tests.

r = .36 (for successes) and .28 (time) with the Vineland feeble-minded group, criterion being Standford-Binet mental age.

Discussion: This test seemed promising at both the upper and lower levels. At the upper end "the test calls for extreme concentration, a large number of the Princeton men spontaneously testifying that the test seemed more fatiguing than other tests" in the Squires performance series. (Squires '26, p. 193)

Reference:

Squires, '26

No. 174 Painted Cube Test

Devised by E. A. Doll.

Date: 1917

Handled by Stoelting.

Material: Model of a three-inch cube, painted red on all surfaces; twenty-seven cubes, eight painted red on three sides, twelve on two sides, six on one side, and one unpainted.

Problem: To build up one large cube, painted all over, like the model.

Individual test; adaptable to group testing.

Directions: Verbal, pantomime by adaptation.

Responses: Manipulative.

Time: Up to twenty minutes.

Scoring: Time.

Standardization: This test was standardized by McFarlane, on 138 English children belonging to a technical school. Norms are expressed in terms of sex and quartiles for ages fourteen and sixteen years only.

Discussion: McFarlane thinks that this test, in common with the Healy Puzzle Box, operates in testing "the power to grasp directly spatial, temporal, and qualitative relations between concrete things." These factors, she considers, go to make up "practical ability." (McFarlane, '24, p. 35)

References:

Doll, '17
Link, '19
McFarlane, '24
Dodd, '26
Bronner, Healy, Lowe, Shimberg, '27

No. 175 Relationship Test

Devised by E. H. Mullan.

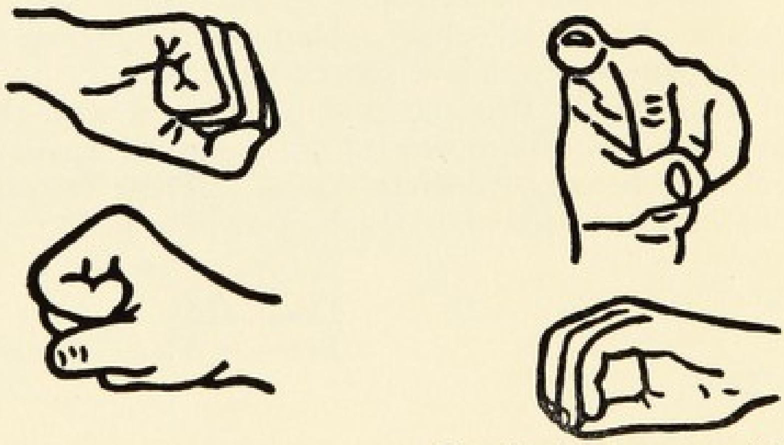
Date: 1917

Material: Two parts, called the "board" and the "field."

1. The board is 18½ x 5¾ inches and contains two rows of designs securely attached, the designs of the upper row being the same in form and color as those immediately beneath them in the lower row, but larger in size. There is one vacant place in each row, corresponding to a square and a triangle, respectively, in the parallel row.
2. The field, situated to right of board, contains thirteen detached designs of various sizes, colors and shapes, among them being a large and a small triangle.

Problem: After the relationship between the corresponding pieces in the upper and lower rows has been pointed out, Subject is required to find the pieces in the field which belong in the vacant places on board.
Individual test; not adapted to group testing.
Directions: Verbal and pantomime.
Responses: Manipulative.
Time: One minute after completion of instructions.
Scoring: One point for correct selection of piece for each vacant place; one-half point for right shape and wrong size.
Standardization: This test is not standardized. Mullan reports results from 292 literate and illiterate aliens at Ellis Island, ranging from nine years to adult. No significant age differences are revealed.
Discussion: This test was suggested to Mullan by R. S. Woodworth as a concrete form of the mixed relations or analogies test. The problem is to grasp the uniform relationship holding between the upper and lower rows and apply this to the filling of the vacant places. "The sagacious subject in doing the test surveys the board, observing its component parts and their relationships." (Mullan, '17, p. 105)
Reference:
Mullan, '17

No. 176 Right and Left Hands Test



Reproduced by courtesy of Princeton Univ. Press

Devised by L. L. Thurstone; modified by P. C. Squires.
Date: 1926
Produced in the Psychological Laboratory, Princeton University.
Materials: Drawings, which depict the right and left hands in various positions, are glued to the face of blank cards; two cardboard boxes. (Squires, '26, p. 132f)
Problem: Given a series of pictures showing the right and left hands, in various positions, to sort all right hands into one box and all left hands into another.
Individual test: Seven practice items and thirty test items. This test is adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.

Time: No time limit.
Scoring: Time for each of five sets, but not for single item; total number of moves; successes minus failures.
Standardization: Squires experimented with this test on fifty Princeton seniors and fifty feeble-minded males at Vineland. Norms have not been established.
Validity: r = .10 (successes) and — .24 (time) for Princeton group, criterion being the higher score on Thorndike or Princeton intelligence examination. r = .16 (successes) and .06 (time) for Vineland group, criterion being Stanford-Binet mental age. (Squires, '26, p. 184)
Discussion: Squires found it easier to "put across" this test with the Princeton than with the Vineland groups, the feeble-minded men finding it especially difficult to grasp the pantomime instructions. (Squires, '26, p. 194)
References:
Squires, '26

No. 177 Single Cube Piles Test

Devised by C. C. Brigham; modified by P. C. Squires.
Date: 1926
Produced by the Psychological Laboratory, Princeton University.
Material: Two boards, A, 30 x 6 x ¼ inches, and B, 27 x 6 x ¼ inches, each fitted with platforms and piles of wooden counters, permanently mounted in front of the platforms: fourteen cube piles, a square wooden plate and a screen, 2 x 1 feet, fourteen wooden pointers, 7 x 8 inches, etc. For details see Squires. ('26, p. 19)
Problem: Given a number of cubes, some of which are concealed by others, to determine the number in the pile.
Individual test: Series A and B; twelve items and two practice items. This test is not adaptable to group testing.
Directions: Pantomime.
Responses: Manipulative.
Time: No time limit.
Scoring: Time; moves; successes.
Standardization: This test is not yet standardized. Squires has tried it out experimentally on fifty Princeton men and fifty feeble-minded men.
Validity: r = .26 (successes) and — .50 (time) for the former group, the criterion being the higher score on either the Princeton or Thorndike intelligence examination. r = .35 (successes) and .14 (time) with the Stanford-Binet mental age for the feeble-minded subjects.
Discussion: Squires reports that this test worked fairly well in both groups, but better for the Princeton than for the Vineland, feeble-minded, men. (Squires, '26, p. 191)

A considerable amount of work on the Squires material, which might be described as preliminary orientation, has been done by Hoffman, a graduate student at Ohio State University. Maxfield writes that the piled cubes were cumbersome as test material, that they frequently came to pieces and required repair, etc. Drawings were then made of these piled cubes on square cards, and Squires' method was adapted to this material. No attempt was made to set up equivalence with the Squires material. The preliminary work established the point that these drawings can be used in place of the actual cubes in a strictly non-verbal test situation. The drawings were similar to those shown in Squires' manual; also similar to those used in the U. S. Army Beta test.

References:

Squires, '26

Maxfield, letter to writers.

No. 178 Spatial Relations

Devised by L. L. Thurstone.

Date: 1911 (?)

Published by Stoelting.

Price: $2.50 per hundred for each form.

Material: Paper on which are twenty-four series of rhomboid figures, in form A, and thirty-nine figures, in form B. In form A, these rhomboid figures are arranged in three columns. Each rhombus in the first column has a small circle drawn in one corner; opposite each rhombus are two other rhombuses without the circle. In form B, are pairs of diagrams, in each of which are two circles; some of these pairs are exact mates, some are not.

Problem: Form A: To superimpose mentally one rhombus upon the other; to find the one of the two which corresponds to the rhombus with the circle, and then to draw the circle in the corresponding rhombus. Form B: To indicate by a plus or minus sign whether the pairs given in each case represent the same face.

Individual or group test.

Directions: Verbal; pantomime adaptation doubtful.

Responses: Non-verbal.

Scoring: Number right.

Standardization: This test is not yet standardized. It was tried out experimentally by Steacy, who used it on thirty-one boys and thirty-eight girls, in the sixth grade of a New York public school.

Reliability: $r = .58$ for one form against another.

Validity: $r = .28$ with the rest of the Steacy battery of tests. (Steacy, '19, p. 41)

Discussion: This is one of twelve tests which Thurstone and W. B. Jones have included in a battery, but for which they have not yet published results.

References:

A. L. Rogers, '18 and '19

Steacy, '19

Payne, '25

Bronner, Healy, Lowe and Shimberg, '27

W. B. Jones, letter to writers.

XIII
GROUP TESTS

No. 179 Army Beta

Devised by Psychology Committee of National Research Council for the United States Army.

Date: 1920

Published by Stoelting; also by Kansas State Teachers College.

Price: $2.25 per twenty-five; $6.75 per hundred. (Stoelting).

Materials: Paper folder containing a series of pictures or drawings; also blackboard frame for demonstration; a Beta chart and six cardboard pieces for Test 7. One form only.

Group test. Scale: 1, maze; 2, cube analysis; 3, X-O series; 4, digit-symbol; 5, number checking; 6, pictorial completion; 7, geometrical construction.

Directions: Verbal or pantomime.

Responses: Non-verbal.

Time: About half an hour.

Scoring: By stencils and key; total score is number of raw scores on separate tests.

Standardization: This test was standardized by the Division of Psychology, S. G. O., U. S. Army, for the purpose of testing illiterates, foreign men, and those who had special verbal handicaps. Number of cases: 23,547. Norms are expressed in terms of raw scores, letter-ratings and percentiles.

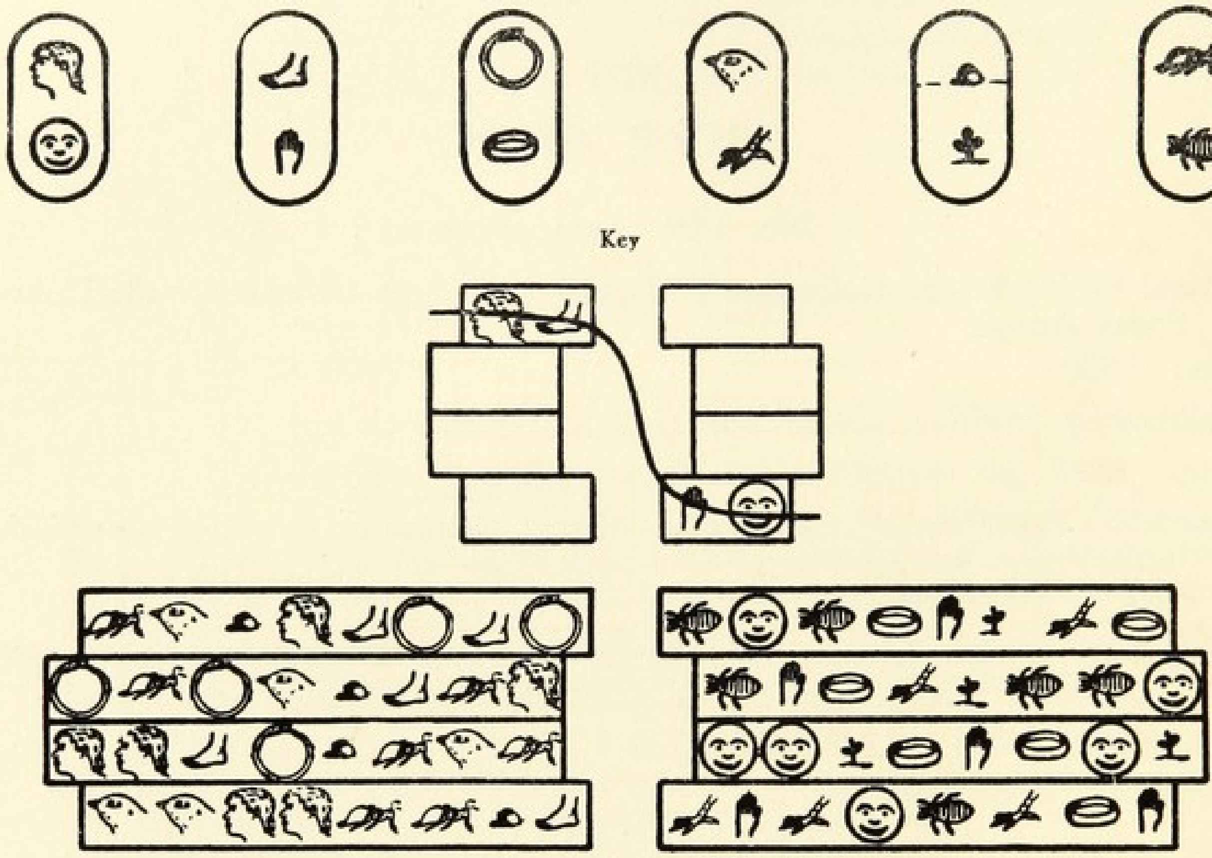
Validity: r = .80 with Army Alpha.
r = .73 with Stanford-Binet.
r = .91 with a composite of Stanford-Binet, Alpha and Beta. (Yoakum and Yerkes, '20, p. 20)

Discussion: This test is particularly adapted to people who cannot speak English, or any other language, or who are unable to read and write.

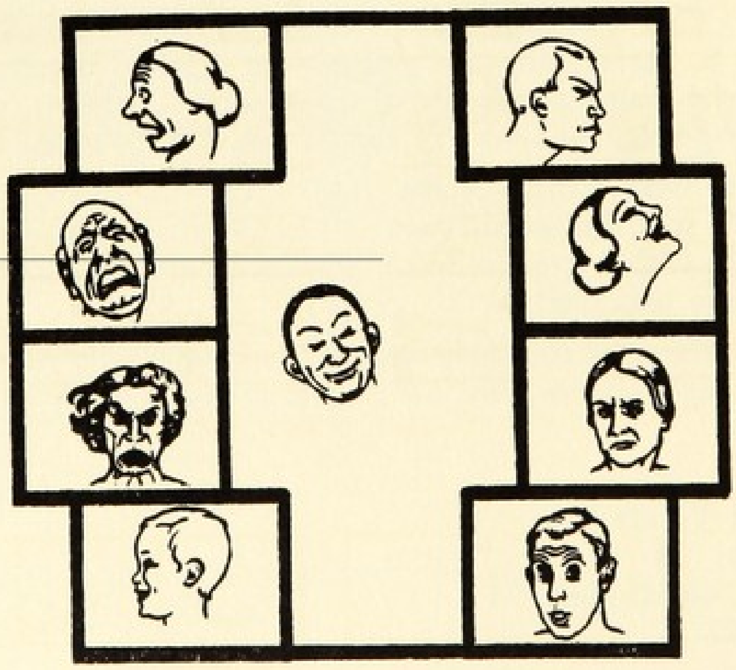
References:

Franz, '19
Yoakum and Yerkes, '20
Yerkes (Ed.), '21
Trabue and Stockbridge, '21
K. Young, '22
Brigham, '23
Pintner, '23
Freeman, '26

No. 180 International Group Mental Tests

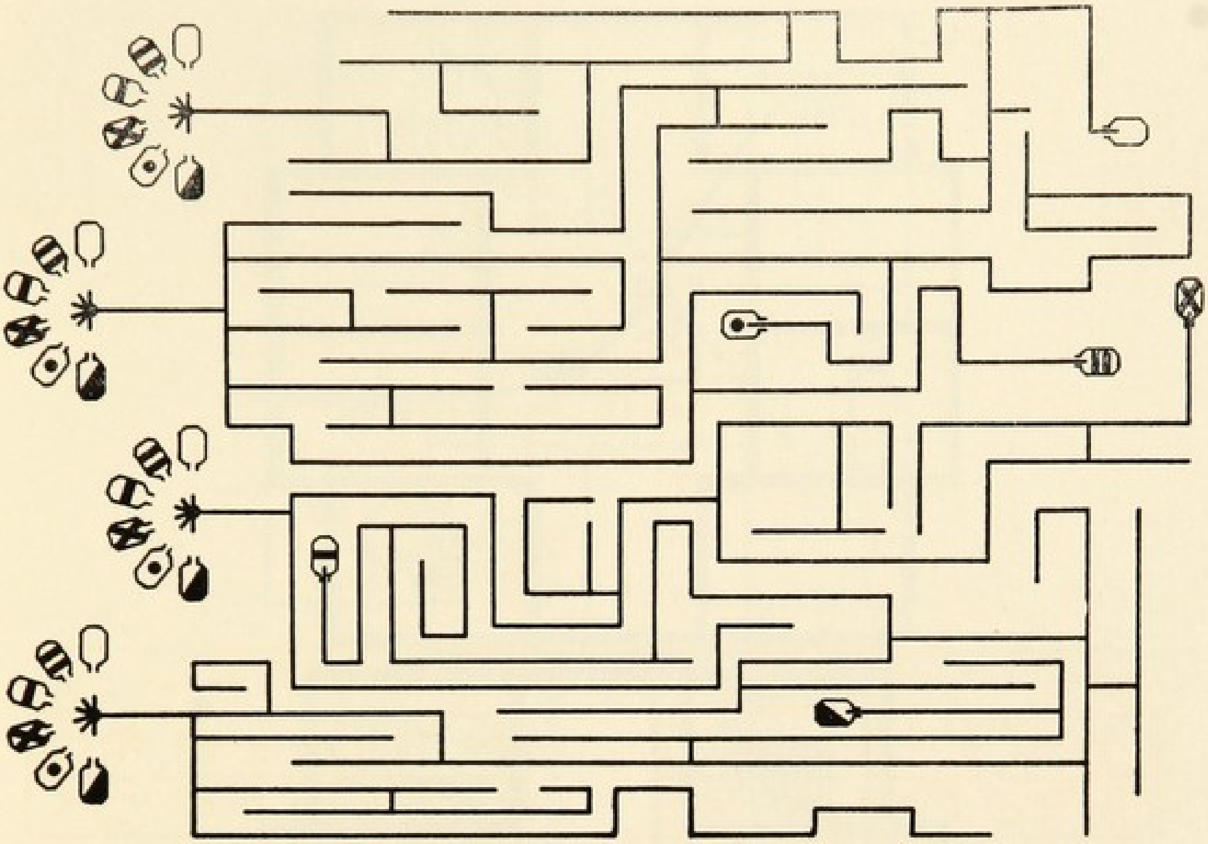


MATCHING ASSOCIATED PICTURES IN INCREASINGLY LARGE GROUPS

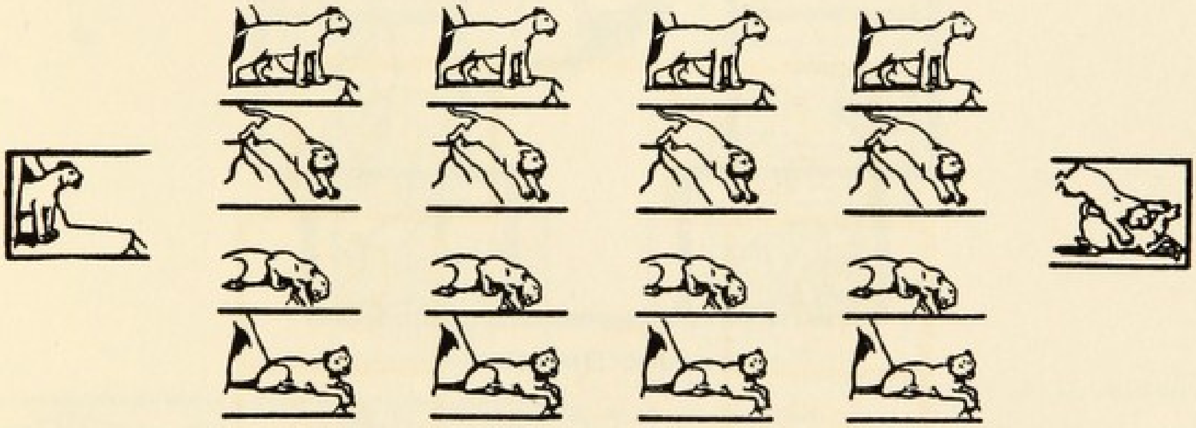


SIMILARITIES IN FACIAL EXPRESSIONS AND EMOTIONS

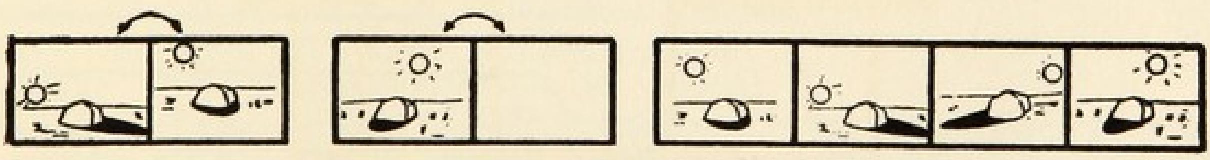
Reproduced by courtesy of C. C. Brigham



MAZES

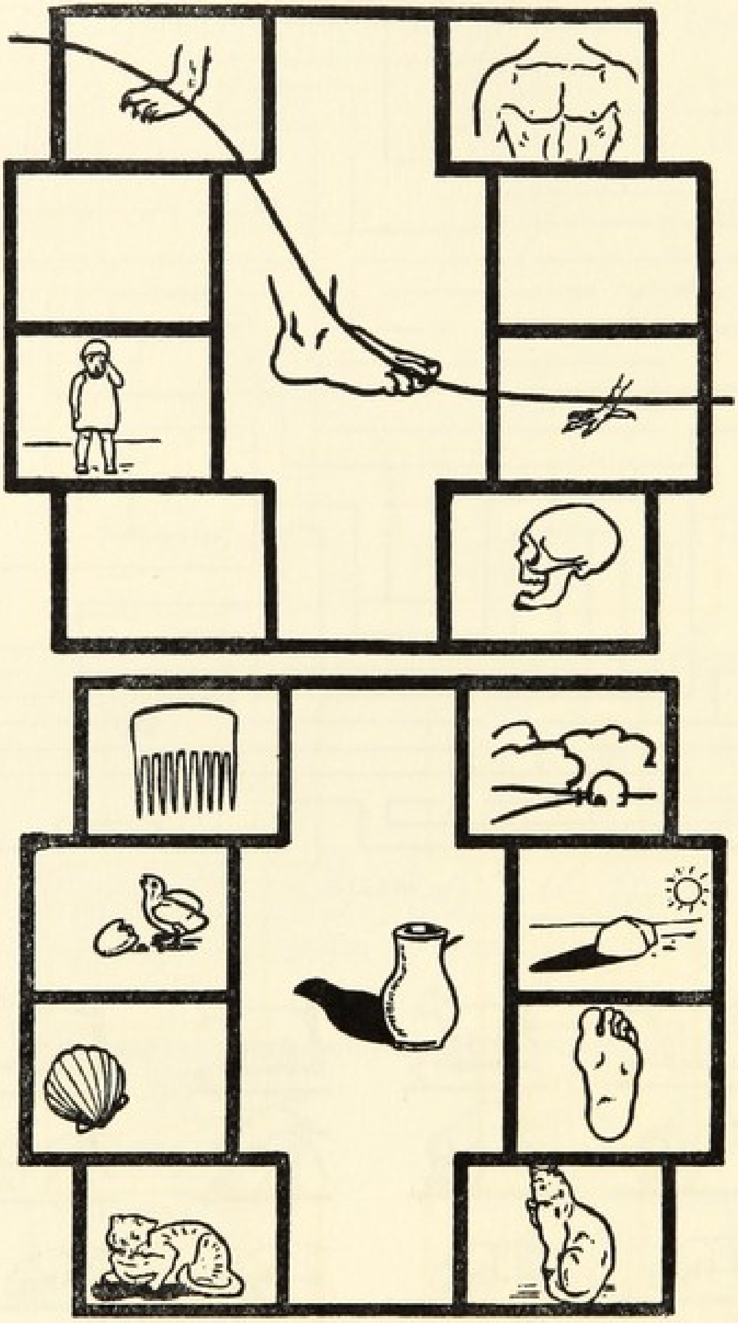


NARRATIVES OR TIME SERIES ARRANGEMENTS

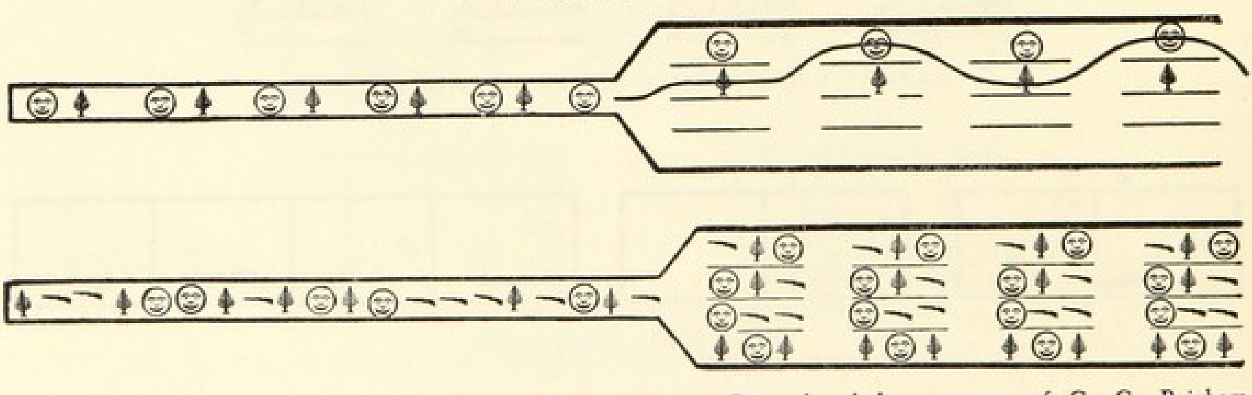


PICTORIAL ANALOGIES

Reproduced by courtesy of C. C. Brigham

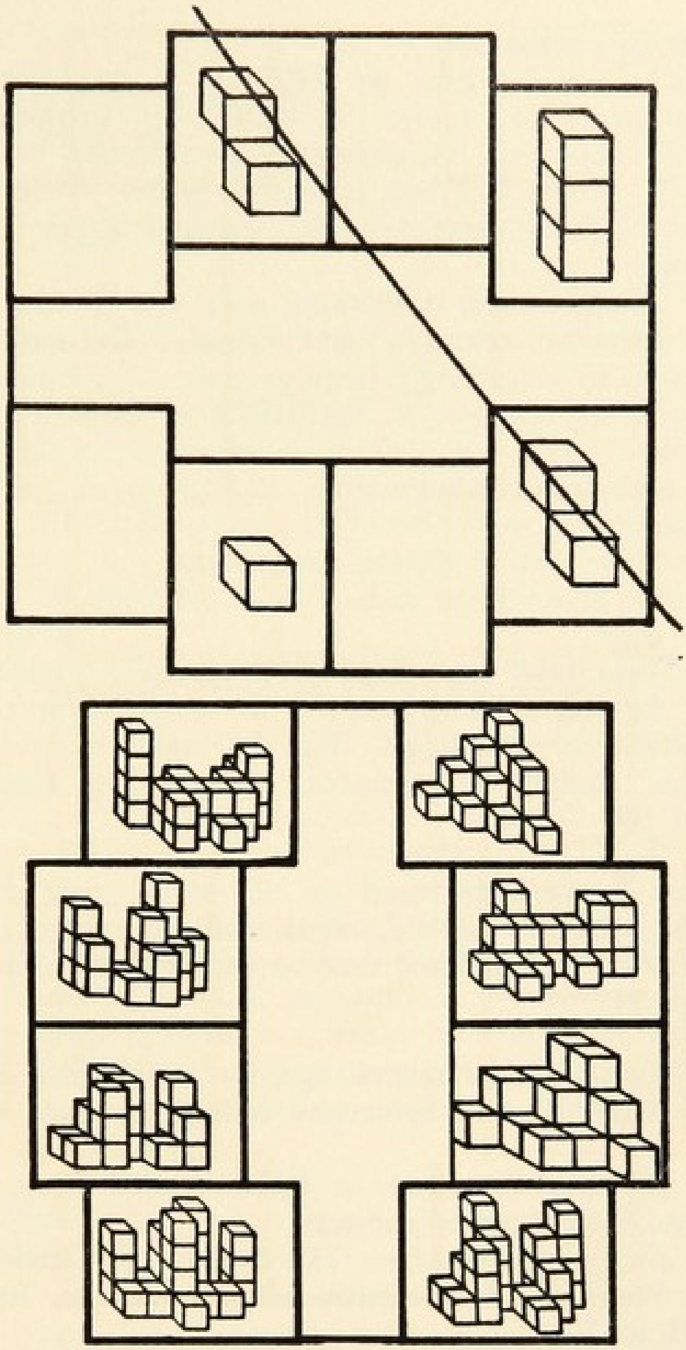


PICTORIAL SIMILARITIES



Reproduced by courtesy of C. C. Brigham

RHYTHM SERIES COMPLETION



Reproduced by courtesy of C. C. Brigham

CUBE COUNTING AND MATCHING

Devised by S. C. Dodd.
Date: 1926
Published by Dodd and Brigham, Princeton University.
Price: See Dodd, '26, Chapter VII and Appendix.

Mean validity: r = .80. (Dodd, '26, p. 3)

Number of forms: Two. Form A: Rotator reaction. Form B: Pencil reaction. (In process of revision).

Materials: Form A: Booklet with rotator devices riveted to the pages, which are made of manila paper. Practice sheets precede each test. A demonstration chart, 28 x 36 inches, has been made for each test to demonstrate clearly the task in pantomime.

Group test: Scale of eleven tests as follows: 1, associations; 2, matching; 3, dots; 4, rhythms; 5, mazes; 6, cubes; 7, similarities; 8, faces; 9, narratives; 10, analogies; 11, geometric form series.

Materials: Form B: Three booklets 10 x 14¼ inches. Practice book; Book 1: tests 1 to 4; Book 2: tests 5 to 8. Revised paper and pencil edition, thirty-eight pages.

Group test. Scale of eight tests, comprising over 400 problems, with over 3500 pictures to be inspected, compared and judged. The tests are: 1, matching associated pictures in increasingly large groups; 2, cube counting and matching; 3, pictorial similarities; 4, similarities in facial expressions and emotions; 5, rhythm series completion; 6, mazes; 7, pictorial analogies; 8, narratives, or time series arrangements.

Directions: Pantomime.

Responses: Non-verbal: rotator device, or cross-out.

Time: Three sittings of an hour each.

Scoring: Number right.

Standardization: These tests have been very carefully prepared and tried out experimentally by Dodd, who considers them applicable to all peoples from three years mental age to college. The tests are not yet usable for purposes of international mental measurement, and further research is necessary. (Dodd, '26, p. 99)

Reliability:

r= .78 (repetition method) based on 112 sixth grade orphans.

r= .97 by the split-half method, on the same group of cases.

r= ..91 for limited vs. unlimited time scores for forty-nine Princeton juniors. (Dodd, '26, p. 3)

Validity:

r= .80 with Stanford-Binet mental age, on 100 feeble-minded subjects.

r= .86 with Thorndike or Princeton college entrance tests on Princeton juniors.

r= .96 with presence in college or feeble-minded home on 149 Princeton juniors and feeble-minded subjects.

r= .71 with chronological age on 283 orphans in Grades II to VI.

r= .74 with Stanford Achievement educational age, for 283 orphans in Grades II to VI.

r= .72 with school grades for 103 orphan twelve-year-olds.

r= .69 with Punjabi Binet M. A. for 147 twelve-year-old Hindu children.

r= .90 with Stanford-Binet mental age for fifty adopted children.

Mean validity: r = .80. (Dodd, '26, p. 3)

Discussion: "In brief the scale appears to have made a successful advance towards universal applicability, to have good reliability and fair validity for a preliminary version, though the criteria are not sufficiently varied to be called 'general intelligence' . . . . All other specifications except that of the cost are reasonably satisfied." (Dodd, '26, p. 4)

"Ability to generalize, to analyze and to understand analogies . . . . is apparently required to the same degree as in verbal tests." (Murdoch, Moddow and Berg, '27)

References:

Dodd, '26
Murdoch, Maddow and Berg, '27
Peterson and Lanier, '29

No. 181 Mental Alertness Test for Illiterates

Devised by W. D. Scott.
Date: 1923
Handled by Stoelting.
Price: \$5.60 per twenty-five; \$16.80 per hundred.
Material: Pencil and paper booklets.
Problem: Six parts. 1, recognition of geometrical objects; 2, missing parts; 3, cancellation; 4, a code test; 5, cube estimation; and 6, numerical sequence.
Group test.
Directions: Verbal or pantomime.
Responses: Non-verbal.
Standardization: These tests are a modification of Army Alpha and Beta. Scott and Clothier devised the series for testing illiterates and people having difficulty with the English language. They report, however, that the tests have not been in general use.
Validity: $r = .51 \pm .0293$ between scores on this test and productivity on the machines. $N = 290$ cases chosen at random from factory employees. (Scott and Clothier, '23, p. 263f)
Discussion: After analyzing their results, Scott and Clothier conclude: "The use of mental tests, although only a partial measurement, is the quickest, most accurate and economical method of prophesying future skill at machines and of placing operators at types of work most suited to their capacity". ('23, p. 266) Wembridge reports much the same data and shows the value of these tests for purposes of prediction where ratings and a manual dexterity test failed.
References:

Scott and Clothier, '23
Wembridge, '23

No. 182 Myers Mental Measure

Devised by G. C. and C. E. Myers.
Date: 1921
Published by Newson and Co.
Price: \$0.10 a copy; \$5.00 a hundred; \$40.00 a thousand. Manual, "Measuring Minds", cloth, pp. 56; price \$0.80.
Material: Paper booklets. Two forms: 1 and 2.
Group test. Scale: 1, following directions; 2, picture completion; 3, first common elements; 4, second common elements.
Directions: Verbal, pantomime by adaptation.
Responses: Non-verbal.

Time: Twenty-five minutes.
Scoring: According to key.
Standardization: This test was standardized by Myers on 15,241 public school children. Norms are expressed as raw scores in terms of age; also "intelligence ratio." Range covered: kindergarten through college.
Validity: r= .80 with Stanford-Binet, based on 300 children evenly distributed through the grades. (Myers and Myers, '21) r= .686 with Stanford-Binet. (Morganthau, '22, p. 45) r = .35 with Army Alpha (Jordan, '23) r= .32 with Miller test of mental ability (Miller, '22) r = .56 with Otis, and r = .55 with Terman Group test, (Terman, '22) r= .49± .03 and r= .43± .03 on 327 foreign and 278 American eleven-year-olds, (V. A. Jones, '27)
Discussion: The test authors show with aid of graphs and tables that the ratings by the Myers Mental Measure "distribute in very close accordance with the probable curve of normal distribution, regardless of the age and school experiences of the groups studied; what some experts have contended could not be done." (Myers and Myers, '21) V. A. Jones says that this test shows no strength in eliminating language factors. He finds it no better than any verbal test in this respect. He adds: "The fact that a test is composed exclusively of non-verbal material is therefore no proof that it has merit as a non-verbal test." ('27, p. 207)
References:

Myers and Myers, '19 and '21
Myers, '20
Doll, '23
Myers and Layton, '21
Liu, '22
Morganthau, '22
Franzen, '22
Miller, '22
Jordan, '23
Sunne, '24
Freeman, 26, (p. 186)
V. A. Jones, '27

No. 183 O'Rourke Non-Language Tests

Devised by L. J. O'Rourke.
Date: 1924
Published by the Education and Personnel Publishing Co., Washington, D. C.
Material: Practice booklet, three pages; test booklet, eight pages. One form. Group test. Scale: 1, Symbol-digit; 2 matching; 3, following lines; 4, digit-symbol; 5, dissecting figures; 6, absurdities; 7, analogies.
Directions: Verbal, pantomime by adaptation.
Responses: Non-verbal.
References:

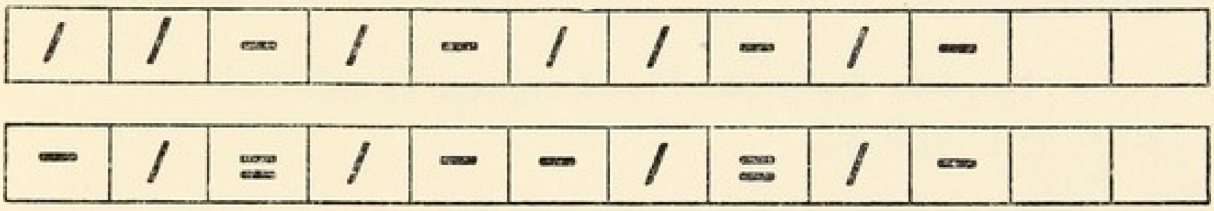
O'Rouke, personal interview.

No. 184 A Pantomime Group Intelligence Test

Devised by G. C. Myers.
Date: 1922
Published by Newson and Co.
Price: $0.15 a copy; $7.50 per hundred; $60.00 per thousand. Demonstrator's edition: $0.40.

Material: Paper booklets. One form.
Group test. Scale: 1, picture completion; 2, similarities; 3, series completion; 4, incongruent elements.

PICTURE COMPLETION



SERIES COMPLETION

INCONGRUENT ELEMENTS

Reproduced by courtesy of Newson & Co.

Directions: Pantomime.
Responses: Non-verbal.
Time: Twenty minutes.
Standardization: This test was standardized by Myers on 1302 public school children. Norms are expressed in terms of raw scores for age and grade. Range: Kindergarten to sixth grade.
References:
Myers, '22
Freeman, '26 (p. 181)

No. 185 Pintner Non-Language Mental Tests

Devised by R. Pintner.
Date: 1920
Published by College Book Company, Columbus, O.
Price: $1.00 per twenty-five; $4.00 per hundred.

Material: Paper booklet of seven pages. One form.
Group test. Scale: 1, imitation; 2, easy learning; 3, hard learning; 4, drawing completions; 5, revised drawing; 6, picture reconstruction.
Directions: Verbal, pantomime by adaptation.
Responses: Non-verbal.
Time: Forty minutes.
Scoring: Points and weighted scores for each test. (Pintner, '19, p. 211)
Standardization: This test was standardized by Pintner on children and adults; 1000 cases for each norm. Norms are expressed in terms of age, grade, percentiles, mental index and standard deviation. Range: Seven years through superior adults.
Reliability: $r = .79 \pm .017$. Re-test coefficient; Grades IV to VI; $N = 201$.
Validity: $r = .78$. Composite criterion; Grades II to IV; $N = 235$. (Liu, '22)
Discussion: It is interesting to note that Pintner was able to administer this test to school children in Belgium, without the aid of language. Comparative results are given by him. (Pintner, '27)
References:
Pintner and Toops, '18
Pintner, '19, '20, '24, '27
Liu, '22
Morganthau, '22
F. N. Freeman, '26

No. 186 Thorndike Non-Language Group Test*

Devised by E. L. Thorndike.
Date: 1919
Formerly published by T. C. Bureau of Publications, Columbia University.
Material: Paper booklets, ten alternative forms.
Group test. Scale: 1, rhythms. 2, dividing geometrical forms. 3, digit-symbol. 4, picture completion. 5, analogies with pictures. 6, mixed spatial relations. 7, memory of objects. 8, easy computation.
Directions: Verbal; pantomime by adaptation.
Time: Part I,, sixteen minutes; Part 1, different series, ten minutes; Part II, sixteen minutes.
Scoring: "Tests I-1, I-3, I-4, and II-3 and II-4 can be scored by any clerk. Tests I-2, II-1, and II-2, require either a moderate amount of judgment, or a few hours of supervision by somebody of good judgment." (Thorndike, '19, p. 24)
Standardization: This test was standardized by Thorndike on 133 adult feeble-minded males, 361 normal sixth grade boys, 169 high school seniors, and 104 superior adults. Norms are expressed in terms of raw scores; also by conversion to Stanford-Binet and Alpha scores. Range: Three years to adult; not suitable for high grade adults.
Reliability: "(P. E. is $\sqrt{2}$). It is therefore probable to estimate the P. E. of a first time score as not over six or seven per cent of the score, for scores from 70 up." (Thorndike, '19, p. 32)

*Although Prof. Thorndike has written that this test is "not for sale, and has not been in use for a number of years" (letter to the writers) data on this test are presented as an example of the kind of information which should be reported for every test.

Validity: r = .84, average of eight correlations; criterion was Stanford-Binet plus schooling, plus five estimates on work reports. (Shepherd's formula). "The authorities of the institution also reported that the rank order of intellect obtained by the test tallied very closely with estimates which they had formed." In the case of the 196 private and non-commissioned officers who were ranked by their officers and fellow-soldiers for intelligence, the correlations were: .23, .33, 43, .50, .53, .58 for the six groups, with a median of .47±.05" . . . "The correlation with the real intelligence possessed by the men would presumably be higher, the amount of intellect imputed by the consensus being itself in each case an imperfect measure of the reality." (Thorndike, '19, p. 25f).

Discussion: The test "seems destined to be useful alone or with other 'performance' tests in the case of immigrants, illiterates and all groups characterized by the absence of opportunity to learn English or any other language." (Thorndike, '19, p. 29).

References:

Thorndike, '19.

PART THREE

BIBLIOGRAPHY OF REFERENCES ON TESTS

Achilles, E. M. '20. Experimental studies in recall and recognition. Arch. of Psychol. No. 44 Pp. 80.

Anderson, L. D. '26. Objective measurements in shop courses. Industrial Arts Magazine. 15: 263-267.

—'27. Environment and mechanical ability. Industrial Psychology. 2, 179-180.

—'28a. The relationship of certain environmental factors to measures of mechanical ability. 27th Yr. Bk. Nat. Soc. Study Educ. Part II, 137-150. Bloomington, Ill., Public School Publishing Co.

—'28b. The Minnesota Mechanical Ability Tests. Person, J. 6, 473-478.

Arthur, M. G. '25. A new point performance scale. J. Appl. Psychol. 9, 390-416.

—'28. The re-standardization of a point performance scale. J. Appl. Psychol. 12: 278-303.

Averill, L. A. '24. Elements of educational psychology. Boston: Houghton Mifflin, Pp. 425.

Baldwin, B. T. '28. The Peg Board as a means of analyzing form perception and motor control in young children. J. Genet Psychol. 35, 389-414.

— and Stecher, L. '25. The psychology of the pre-school child. New York: Appleton, Pp. 305.

Biddelle, A. E. '28. An analytical study of one class in high school. Psychol Clin. 17, 97-118.

Binet, A. '98. La Mésure en psychologie individuelle. Rev. Philos. 46, 115-123.

—'11. La Mésure du développement de l'intelligence chez les jeunes enfants. Soc. libre étude psychol, de l' enfant, 11, 187-248.

—et Simon, Th. '08. Le développement de l'intelligence chez les enfants. Année psychologique. 14, 1-94.

Blacking, E. '27. Standardization of a bead stringing test. Ped. Sem. 34, 620-633.

Blackwood, B. '27. A study of mental testing in relation to Anthropology. Mental Meas. Monog. No. 4 Baltimore: Williams & Wilkins, Pp. 117.

Boder, D. P. '30. A tri-dimensional maze. Amer. Jour. of Psychol. 42, 107-108.

Boody, B. M. '26. A psychological study of immigrant children at Ellis Island. Ment. Meas. Monog. No. 3, Pp. 163.

Bourdon, B. '95. Observations comparative sur la réconnaissance, la descrimination et l'association. Rev. Philos. 40, 153-185.

Bowler, A. C. '17. A picture arrangement test. Psychol. Clin. 11, 37-54.

Brennan, F. M. '26. The relation between musical capacity and performance. Psychol. Monog. 36, Pp. 190-248.

Briggs, T. H. '13. Formal English grammar as a discipline. Teach. Coll. Rec. 14, No. 4. 1-93.

Brigham, C. C. '17. Two studies in mental tests. Psychol. Monog. 24, Pp. 254.

—'23. A study of American intelligence, Princeton, N. J. Princeton Univ. Press, Pp. 210.

—'30. Intelligence tests of immigrant groups. Psychol. Rev. 37, 158-165.

Bronner, A. F. '21. Apperceptive abilities. Psychol. Rev. 28, 270-279.

—'22. The apperceptive abilities of delinquents. J. Delinq. 7, 43-54.

—, Healy, W., Lowe, A. M. and Shimberg, M. E. '27, A manual of individual mental tests and testing. Boston: Little, Brown, Pp. 287.

Brooke, M. C. '28. Results of examination of 91 girls at a State institution. Psychol. Clin., 17, 22-28.

Broom, M. E. '29. The Searshore measures of musical talent. School & Soc. 30, 274-275.

Brotemarkle, R. A. '27. College student personnel problems and individual mental testing at the college-adult level. J. Appl. Psychol. 11, 415-436.

—'28. College student personnel problems II. The analytical study of the student personnel problems. J. Appl. Psychol. 12, 1-43.

Brown, A. J. '30. An enquiry into the standardization of the Kohs Block—Design Test. J. Appl. Psychol. 14, 178-181.

Brown, A. W. '28. The reliability and validity of the Seashore tests of musical talent. J. Appl. Psych. 12, 468-476.

Bruckner, L. and King, L. '16. A study of the Fernald formboard. Psychol, Clin. 9, 249-257.

Burt, C. L. '09. Experimental tests of general intelligence. Brit. J. Psychol. 3, 94-177.

—'19. The development of reasoning in school children. J. Exper. Ped. 5, 68-77. 121-127.

—'23. Handbook of tests for use in schools. London: King & Son. Pp. 106.

—'27. Mental and scholastic tests. London: King & Son. Pp. 447.

Carlisle, C. L. '18. Performance norms for thirteen tests and mental examinations. J. Educ. Psychol. 9, 518-523.

Carter, D. '28. The organization of mechanical intelligence. Ped. Sem. and Gen. Psychol. 35, 270-285.

Clinton, R. J. '30. Nature of mirror drawing ability. Norms on mirror drawing for white children by age and sex. J. Educ. Psy. 21, 221-228.

Cooke, F. '18. A note on mental tests for normal boys. J. Appl. Psychol. 2, 378-380.

Cornell, C. B. '17. A graduated scale for determining mental age. J. Educ. Psychol. 8, 539-550.

Cornell, E. L. '24. Comparison of the Stanford and Porteus tests in several types of social inadequacy. J. Abn. Psychol and Soc. Psychol. No. 1, 33-43.

Cox, C. M. '28a. Comparative behavior in solving a series of maze problems of varying difficulty. J. Exper. Psychol. 11, 202-218.

—'28b. The intelligence factor in the solution of space problems with the two-story maze. Amer. J. Psychol. 40, 542-561.

Dashiell, J. F. '28. Fundamentals of objective psychology. New York: Houghton Mifflin, Pp. 588.

— and Glenn, W. D. '25. A re-examination of a socially composite group with both Binet and performance tests. J. Educ. Psychol. 16, 335-342.

Davenport, C. B. '28. Race crossing in Jamaica. Scien. Mo. 27, 225-238.

—'29. Do races differ in mental capacity? Human Biology. 1, 70-89.

— Unpublished manuscript. "Guide to genetical and sensory mental measurements."

Davey, C. M. '26. A comparison of group verbal and pictorial tests of intelligence. Br. J. Psychol (Gen. Sect.) 17, 27-48.

Dearborn, W. F., Anderson, J. E. and Christiansen, A. O. '16. Construction tests of mental ability. J. Educ. Psychol. 7, 445-458.

— and Brewer, J. M. '18. Methods and results of a class experiment in learning. J. Educ. Psychol. 9, 63-82.

—, Shaw, E. A. and Lincoln, E. A. '23. A series of formboard and performance tests of intelligence. Harvard Monog. Educ. 1, No. 4.

de Fursac, J. R. '11. Manual of psychiatry. (p. 437f) New York: Wiley.

de Sanctis, S. '10. Mental development and measurement of the level of intelligence. Educ. Psychol. 2, 498-507.

Dewey, E., Child, E., Ruml, B. '20. Methods and results of testing school children. New York: Dutton, Pp. 170.

Dodd, S. C. '26a. A statement concerning research on an intelligence scale for international use. Princeton: Princeton Univ. Store, Pp. 19.

—'26b. International group mental tests. Princeton, N. J. Mimeographed only.

Doll, E. A. '13. The DeMoor size-weight illusion. Tr. School Bull. 9, 145-149.

—'17. The painted cube construction test. J. Educ. Psychol. 8, 176-178.

—'23. Nature and purpose of the New Jersey composite test. Tr. School Bull. 20, 33-39.

—'26. Three group tests for illiterates. Tr. School Bull. 22, 122-129.

Dorcus, M. D. '28. Analysis of specific responses in the Healy Picture Completion Test II. J. Genet, Psychol. 35, 574-586.

Dougherty, F. D. '26. A study of the mechanical ability of delinquent children of the Los Angeles juvenile court. J. Delinq. 10, 293-311.

Drever, J. and Collins, M. '28. Performance tests of intelligence. Edinburgh: Olive & Bays, Pp. 52.

Dunham, F. L. '16. Arrow board—an adult formboard test. Ped. Sem. 283-290.

Dunlap, Knight, '22. Improved forms of steadiness testing and tapping plate. Psychol. Rev. 29, 430-433.

Easby-Graves, C. '24. Tests and norms at the six-year-old performance level. Psychol. Clin., 15, 261-300.

Edgerton, H. A. and Paterson, D. G. '26. Table of standard errors and probable errors of percentages for varying number of cases. J. Appl. Psychol., 10, 378-391.

Ellis, F. W. and Bingham, A. T. '15. Seventh Annual Report of New York Probation and Protective Association.

Farnsworth, P. R. '28. The effects of nature and nurture on musicality. The twenty-seventh yearbook National Soc. Stud. Educ. Pt. II, 233-245.

— and Roberts, W. A. '26. An attentional learning board. J. of Educ. Psychol. 17, 275-277.

Farson, M. R. '28. A report on the examination of one hundred VI-B children in Philadelphia schools. Psychol. Clin., 17, 128-152.

Ferguson, G. O., Jr. '20. A series of form boards. J. Exper. Psychol., 3, 47-58.

Fernald, G. G. '12. The defective delinquent class differentiating tests. Amer. J. of Insan., 68, 523-594.

Fernberger, S. W. '27. The application of the Hutt color cube test to a group of subnormal mentality. Psychol. Clin., 16, 232-233.

Fildes, L. G. '25. Performance tests with defectives. Mental Welfare. 6, 88-93.

Foster, W. D. '23. Experiments in Psychology. New York: Holt. Pp. 309-332.

— and Tinker, M. A. '29. Experiments in psychology. (Rev. Ed.) New York: Holt. Pp. 392.

Franz, S. I. '12. Handbook of mental examination methods. Nerv. and Ment. Dis. Monog. Ser. No. 19. Pp. 165.

—'19. Handbook of mental examination methods. New York: Macmillan. Pp. 193.

Franzen, R. H. '22. Attempts at test validation. J. Educ. Res., 6, 145-158.

Freeman, F. N. '16. Experimental education. Boston: Houghton-Mifflin.

—'26. Mental tests. Boston: Houghton-Mifflin.

Garrett, H. E. and Lemmon, V. W. '24. An analysis of several well-known tests. J. of Appl. Psychol. 8, 424-438.

Gault, R. H. '20. Picture Completion. J. Appl. Psychol. 4, 310-315.

Gaw, F. '25a. A study of performance tests. Brit J. of Psychol. 15, 374-392.

Gaw, F. '25b. Performance tests of intelligence. Report No. 31. Industrial Fatigue Research Board. London: H. M. Stationery Office.

Goddard, H. H. '08. The Binet and Simon tests of intellectual capacity. Tr. School. 5.

—'11. A revision of the Binet scale. Tr. School. 8, 56-62.

—'12. The adaptation board. Proc. of the Wash. meeting of the Psychol. Association. Psychol. Bulletin. 9, 79.

—'26. The formboard as a measure of intellectual development in children. Tr. School. 9, 49-52.

—'12. The feeble-minded immigrant. Tr. School Bull. 9, 109-113.

—'15. The adaptation board as a measure of intelligence. Tr. School. Bull. 11, 182-188.

—'17. Mental tests and the immigrant. J. Delinq., 2, 243-277.

Goodenough, F. L. '26. Measurement of intelligence by drawings. Yonkers: World Book Co. Pp. 177.

—'26. A new approach to the measurement of the intelligence of young children. Ped. Sem. 33, 185-211.

—'26. Racial differences in the intelligence of school children. J. Exper. Psychol. 9, 388-397.

—'27. The reliability and validity of the Wallin peg boards. Psychol. Clin. 20, 199-215.

Gopalaswami, M. V. '24. Intelligence in motor learning. Brit. J. of Psychol. 14.

Goudge, M. E. '15. A simplified method of conducting McDougall's spot pattern test. J. Educ. Psychol. 6, 73-84.

Graham, J. '26. The intelligence of Italian and Jewish children. J. Abn. and Soc. Psychol. 20, 371-376.

Graham, J. L. '30. A quantitative comparison of certain mental traits of Negro and white college students. J. Soc. Psychol, 1.

Gwin, M. K. '14. The puzzle picture and defective aliens. Med. Record. 85, 197-199.

Haines, T. H. '19. Mississippi mental deficiency survey. Nat. Comm. for Ment. Hygiene. Pp. 45.

Hallowell, D. K. '28. Mental tests for pre-school children. Psychol. Clin. 16, 235-276.

Healy, W. '14. A pictorial completion test. Psychol. Rev. 21, 189-203.
—'15. The individual delinquent—a text-book. Boston: Little Brown and Co. (Pp. 86 and 106).
—'21. Pictorial completion test II. J. Appl. Psychol. 5, 224–239.
— and Fernald, G. '11. Tests for practical mental classification. Psychol. Monog. 13. Pp. 54.

Hewes, A. '22. Standardization of the tapping test. J. Apply Psychol. 6, 113–119.

Hollingworth, H. L. '20. The psychology of functional neurosis. New York: Appleton. Pp. 259.

Hollingworth, L. S. and Monahan, J. E. '26. Tapping rate of children who test above 135 I. Q. (Stanford-Binet). J. Educ. Psychol. 17, 505–518.

Hubbard, R. M. '28. A measurement of mechanical interests. Ped. Sem. and J. of Gen. Psychol. 35, 229–254.

Hunter, W. S. '23. General Psychology. (Rev. Ed.) Chicago: Chicago Press. Pp. 368.

Husband, R. W. '28. Human learning on a four-section elevated finger maze. J. Gen. Psychol. 1, 15–28.

Hutt, R. B. W. '25. Standardization of a color cube test. Psychol. Clin. 16, 77–97.

Ide, G. G. '18. The Witmer formboard and cylinders as tests for children two to six years of age. Psychol. Clin. 12, 65–88.

Jarrett, R. F. '26. Some observations on social capacity; application of the Porteus maze tests to 100 Borstal lads. Lancet, 1926, 2, 1059–1060.

Jastrow, J. '98. A sorting apparatus for the study of reaction time. Psychol Rev. 5, 279–285.

Johnson, B. J. '25. Mental growth of children. New York: Dutton.
— and Schreifer, L. '22. A comparison of mental age scores obtained by performance tests and the Stanford revision of the Binet-Simon scale. J. Educ. Psychol. 8, 408–418.

Jones, A. M. '25. An analytical study of one hundred twenty superior children. Psychol Clin. 16, 19–76.

Jones, V. A. '27. A study of the non-verbal nature and validity of Myers' Mental Measure. J. Educ. Res. 16, 203–209.

Jordan, A. M. '23. The validation of the intelligence tests. J. of Educ. Psychol. 14, 414–428.

Keane, F. L. and O'Connor, J. '27. A measure of mechanical aptitude. Personn. J. 6, 15–24.

Kefauver, G. N. '29. Relationship of the intelligence quotient and scores on mechanical tests with success in industrial subjects. Voc. Guid. Mag. 7, 198–203.

Kelley, T. L. '16. A constructive ability test. J. Educ. Psychol. 7, 1–17.

Kent, G. H. '16. A graded series of colored picture puzzles. J. of Exper. Psychol. 1, 242–246.
—'16. A graded series of geometrical puzzles. J. of Exper. Psychol. 1, 40–50.
—'23. A combination mental test for clinical use. J. of Appl. Psychol. 7, 246–257.
— and Shakow, D. '28. A graded series of form boards. Person. J. 7, 115–120.

Kline, L. W. and Kline, F. L. '27. Psychology by experiment. Boston: Ginn and Co. p. 216.

Klineberg, O. '27. Racial differences in speed and accuracy. J. Abn. and Soc. Psychol. 22, 273–277.

Knox, H. G. '13a. A test for adult imbeciles and six-year-old normals. New York, Med. J. 98, 1017–1018.
—'13b. Two new tests for detection of defectives. N. Y. Med. J. pp. 522.
—'14. A scale based on the work at Ellis Island for estimating mental defect. J. Amer. Med. Assoc. 62, 741–747.

Kohs, S. C. '20. The block design tests. J. Exper. Psychol. 3, 357–376.
—'23. Intelligence measurement: A psychological and statistical study based upon the Block Design tests. New York; Macmillan. Pp. 312.

Kuhlmann, F. '12. A revision of the Binet-Simon system for measuring the intelligence of children. J. Psycho-Asthenics. Monog. Suppl. No. 1. Pp. 41.
—'22. Handbook of mental tests. Baltimore, Md.; Warwick and York.

Kwalwasser, J. '27. Tests and measurements in music. Boston; Birchard. Pp. 159.
—'28. Tests and measurements in music. Psychol. Bull. 25, 284–301.

Langfeld, H. S. and Allport, F. H. '16. An elementary laboratory course in psychology. Boston; Houghton-Mifflin. Pp. 147.

Lanier, L. H. '27. Prediction of the reliability of mental tests and tests of special abilities. J. Exper. Psychol. 10, 69–113.

Larson, D. L. '28. An experimental critique of the Seashore consonance test. Psychol. Monog. No. 176. 38, 49–81.

Leaming, R. E. '22. Tests and norms for vocational guidance at fifteen year old level. Psychol. Clin. 14, 193–220.

Lester, O. P. '29. Performance tests and foreign children. J. Educ. Psychol. 20, 303–309.

Lincoln, E. A. '27. Tentative standards for the Lincoln hollow square formboard. J. Appl. Psychol. 11, 264–267.

Link, H. C. '19. Employment psychology. New York; Macmillan. Pp. 440.

Liu, H. C. '22. Non-verbal intelligence tests for use in China. T. C. Col. Contrib. Educ. No. 126.

Lough, J. E. '12. Plateau in simple learning. Psychol. Bull. 9, 87–88.

Lowe, G. M., Shimberg, M. S. and Wood, M. W. '24. Further standardization of construction tests A and B. J. Appl. Psychol. 8, 324–338.

MacPhee, E. D. and Brown, A. J. '30. An enquiry into the standardization of the Ferguson form boards. J. Educ. Psychol. 21, 24–36.

MacQuarrie, T. W. '27. A mechanical ability test. J. Personn. Res. 5, 329–337.

Mateer, F. '24. The unstable child. New York; Appleton, Pp. 471.

Maxfield, F. N. '25. Design blocks. Psychol. Clinic. 16, 98–109.

McFarlane, M. M. '24. A study of practical ability. Brit. J. Psychol. Monog. Suppl. No. 8. 1–75.

McGinnis, E. '28. Seashore's measures of musical ability applied to children of pre-school age. Amer. J. Psychol. 40, 620–623.

—'29. The acquisition and interference of motor habits in young children. Gen. Psychol. Monog. 6, No. 23.

—'28. A child's stylus maze. Amer. J. Psychol. 40, 313.

Miles, W. R. '27. The two-story duplicate maze. J. Exper. Psychol. 10, 365–377.

—'28. The high relief finger maze for human learning. J. of Gen. Psychol. 1, 3–14.

Miller, W. S. '22. The administrative use of intelligence tests in the high school. 21st. Yrbk. Nat. Soc. Study Educ. Part II, 189–222.

Mohr, G. J. and Gundlach, R. H. '27. The relation between physique and performance. J. Exper. Psychol. 10, 117–157.

Morganthau, D. R. '22. Some well known mental tests evaluated and compared. Arch. Psychol. 7, 5–54.

Mulhall, E. '17. Tests of the memories of school children. J. of Educ. Psychol. 1, 294–302.

Mullan, E. H. '17. The mentality of the arriving immigrant. Publ. Health Bull. No. 90. Washington, D. C., Gov't Print. Office. Pp. 132.

Murdoch, K. '25. A study of the comparative value of nine performance tests. J. Appl. Psychol. 9, 364–366.

Murdoch, K., Maddow, D. and Berg, N. L. '27. A study of the relation between intelligence and the acquisition of English. 27th Yrbk. Nat. Soc. Study. of Educ. Part I, 343–353.

Murphy, M. '28. The ten year level of competency. Psychol. Clin. 17, 33–60.

Myers, G. C. '20. A grave fallacy in intelligence test correlation. School and Soc. 11, 528–529.

—'22. Pantomime group intelligence test. New York: Newson and Co.

— and Myers, C. E. '19. A group intelligence test. School and Soc. 10, 355–366.

—'20. The Myers mental measure. The Sentinel. Carlisle, Pa.

— Myers Mental Measure. 21st Yrbk. of Nat. Soc. Educ., Chap. IV. Public School Pub. Co.

— Myers Mental Measure and measuring minds (examiner's manual). New York; Newson.

— and Layton, S. H. '21. Group mental testing in Altoona, Pennsylvania. School and Soc. 13, 624–628.

New York Bureau of Analysis and Investigation. '15. Eleven mental tests standardized. Albany, New York; Eugen. and Soc. Welfare Bull. No. 5.

—'17. Performance norms for thirteen tests. Albany, New York; Eugen. and Soc. Welfare Bull. No. 8.

Norsworthy, N. '06. The psychology of mentally deficient children. Arch. Psycol. 1, Pp. 111.

Nute, Dr. '15. Public Health Service Report. No. 15.

O'Connor, J. '28. Born that way. Baltimore; Williams and Wilkins. Pp. 296.

O'Rourke, L. J. '24. O'Rourke non-language tests. Washington, D. C.: Education and Personnel Publ. Co.

Paschal, F. C. '18a. A report on the standardization of the Witmer cylinder test. Psychol. Clin. 12, 54–59.

—'18b. The Witmer cylinder test, Hershey, Pa; The Hershey Press. Pp. 89.

Paterson, Donald G. '28. The Minnesota mechanical ability tests. Vocational Guidance Bulletin. Minneapolis Public Schools, 2, No. 9., 1–2.

—, Elliott, R. M. Anderson, L. D., Toops, H. A., and Heidreder, Edna. Minnesota mechanical ability tests. To be published by the University of Minnesota Press. About 500 p.

Payne, A. F. '25. Organization of vocational guidance. New York: McGraw-Hill. Pp. 438.

Perrin, F. A. C. '19. The learning curves of the analogies and the mirror reading tests. Psychol. Rev. 26, 42–62.

— and Klein, D. B. '26. Psychology: its methods and principles. New York; Holt. Pp. 387.

Perry, D. E. '22. Interpretations of the reactions of the feeble-minded on Healy Pictorial Completion Test II; social implications. J. Delinq. 7, 75–85.

Peterson, John C. '20. The higher mental processes in learning. Psychol. Rev. Monog., 28, Pp. 121.

Peterson, Joseph, '18. Experiments in rational learning. Psychol. Rev., 25, 443–467.

—'20a. The rational learning test applied to eighty-one college students. J. Educ. Psychol., 11, 137–150.

—'20b. Tentative norms in the rational learning test. J. Appl. Psychol., 4, 250–257.

—'22. Tentative norms of a simplified rational learning test for children, eight, nine, and ten years of age. Va. Teach., 3, No's. 9–10.

—'23. Comparative abilities of white and Negro children. Comp. Psychol. Monog., 1, Pp. 141.

—'23. Comparison of white and Negro children in rational learning test. 27th Yrbk. Nat. Soc. Stud. Educ., Pt. 1, 333–341.

— and Lanier, L. H. '29. Studies in the comparative abilities of whites and Negroes. Ment. Meas. Monog. No. 5. Baltimore, Md.; Williams and Wilkins, Pp. 156.

— and Walker, H. M., '25. Comparisons of white and Negro children in certain ingenuity and speed tests. J. Comp. Psychol., 5, 271–284.

Pintner, R., '15. The standardization of Knox's cube test. Psychol. Rev., 22, 377–401.

—'18. The mental survey. New York: Appleton. Pp. 116.

—'19. A non-language group intelligence test. J. Appl. Psychol., 3, 199–214.

—'20. Manual of directions for the non-language mental and educational survey tests. Columbus, Ohio; College Book Co., Pp. 16.

—'23. Comparison of American and foreign children on intelligence tests. J. Educ. Psychol., 14, 292–295.

—'24. Results obtained with the non-language group test. J. Educ. Psychol. 15, 473–483.

—'27. Non-language tests in foreign countries. School and Soc., 26, 374–376.

—'28. A mental survey of the deaf. J. Educ. Psychol. 19. 145–151.

— and Anderson, M. M. '17. The picture completion test. Educ. Psychol. Monog. Pp. 101.

— and Marshall, H. '21. A combined mental educational survey. J. Educ. Psychol. 12, 32–43.

— and Paterson, D. G. '15. The Binet scale and the deaf child. J. Educ. Psychol. 6, 201–210.

—'15. The factor of experience in intelligence testing. Psychol. Clin. 9, 44–50.

—'16a. The formboard ability of young deaf and hearing children. Psychol. Clin. 9, 234–237.

—'16b. Learning tests with deaf children. Psychol. Rev. Monog. 20, No. 88.

—'17. A scale of performance tests. New York; Appleton, Pp. 218.

— and Toops, H. A. '18. A drawing completion test. J. Appl. Psychol. 2, 163–173.

Poffenberger, A. T. '16. Experimental psychology. Loose leaf laboratory manual. New York; Morningside Press. Pp. 50.

Porteus, S. D. '15a. Motor intellectual tests for defectives. J. Exper. Ped. 3, 127–135.

—'15b. Mental tests for the feeble-minded: a new series. J. Psycho-Asthenics. 19, 200–213.

—'17. Mental tests with delinquents and Australian aboriginal children. Psychol. Rev. 24, 32–41.
—'18. Measurement of intelligence: 653 children examined by the Binet and Porteus tests. J. Educ. Psychol. 9, 13–31.
—'19. Porteus test. The Vineland Revision. Tr. School. Publ. Dept. Res. No. 6. Pp. 44.
—'22. Studies in mental deviations. Tr. School Res. Publ. No. 24. Pp. 276.
—'24a. Temperament and mentality in maturity, sex and race. J. Appl. Psychol. 8, 57–74.
—'24b. Porteus tests and social inadequacy. J. Abn. and Soc. Psychol. 19.
—'25. Guide to the Porteus maze tests. Tr. School. Res. Bull. No. 25. Pp. 50.
— and Babcock, M. E. '26. Temperament and race. Boston: Badger. Pp. 364.
— and Berry, R. J. A. '20. Intelligence and social valuation. Tr. School Res. Publ. No. 20. Pp. 100.
— and Hill, H. F. '20. Condensed guide to the Binet tests. Tr. School Publ. Dept. No. 20.
Poull, L. E. and Montgomery, R. '29. The Porteus Maze Test as a discriminative measure in delinquency. J. Appl. Psychol. 13.
Proceedings and papers. '30. Ninth International Congress in Psychology. Princeton: Psychol Rev. Pp. 534.
Pyle, W. H. '13. The examination of school children. New York: Macmillan. Pp. 70.
—'15. The mind of the Negro child. School and Soc. 1, 357.
—'16. Manual for the mental and physical examination of school children. Bull., Univ. of Mo., 17, No. 24. Pp. 32.
—'19. Is individual learning capacity constant for different types of material? J. Educ. Psychol. 10, 121–128.
—'23. A laboratory manual in the psychology of learning. Baltimore, Md.: Warwick and York. Pp. 161.
—'25. The nature and development of learning capacity. Baltimore, Md.: Warwick and York. Pp. 122.
—'28. The psychology of learning. (Rev. Ed.) Baltimore, Md.: Warwick and York. Pp. 441.
Rachofsky, L. M. '18. Speed and presentation and ease of recall in Knox Cube Tests. Psychol. Bull. 15, 61-64.
Riley, G. '29. Stanford-Binet 'indicators' of mechanical ability. Psychol. Clin. 18, 128–132.
Rogers, A. L. '18. Experimental tests of mathematical ability and their prognostic value. Teach. Coll. Contrib. Educ. No. 89, Pp. 118.
—'19. Tests of mathematical ability: their scope and significance. Math. Teacher. 6, 145–164.
Rogers, H. W. '24. Research on mechanical ability. Voc. Guid. Mag. 11, 178–180.
Ruger, H. A. '10. The psychology of efficiency. Arch. Psychol. 2, Pp. 88.
Schmitt, C. '15. Standardization of tests for defective children. Psychol. Monog. 19, Pp. 179.
Schoen, M. '23. The validity of tests of musical talent. J. Comp. Psychol. 3, 101–121.
Scott, W. D. and Clothier, R. '23. Personnel management, (pp. 262–266). Chicago and New York: Shaw. Pp. 643.
Seashore, C. E. '08. Elementary experiments in psychology. New York: Holt. Pp. 218.
—'19. The psychology of musical talent. Boston: Silver, Burdett. Pp. 288.
—'30. Measures of musical talent. A reply to Dr. C. P. Heinlein. Psychol. Rev. 37, 178–183.
Shakow, D. and Kent, G. H. '25. The Worcester formboard series. Ped. Sem. 32, 599–611.
Shaw, E. A. '18. A new picture completion test. J. Appl. Psychol. 2, 355–365.
Sherman, I. C. '23. The Franz dot tapping test as a measure of attention. J. Appl. Psychol. 7, 353–359.
Simpson, B. R. '12. Correlations of mental abilities. Teach. Coll. Contrib. to Educ. Pp. 122.
Skaggs, E. B. '20. Comparison of results obtained by the Terman-Binet tests and the Healy picture completion test. J. Educ. Psychol. 11, 418–421.
Smith, S. '27. A scale of individual tests. Seattle: Univ. of Wash. Publ. In Soc. Sc., 2, 183–204.
Smith, H. L. and Wright, W. W. '28. Tests and measurements. Boston: Silver, Burdett. Pp. 540.
Snoddy, G. S. '20. An experimental investigation of trial and error learning in human subjects. Psychol. Monog. 28, Pp. 57.

—'26. Learning and stability. J. Appl. Psychol. 10, 1–36.

Squires, P. C. '26. A universal scale of individual performance tests. Princeton, Princeton Univ. Press. Pp. 158.

Stanton, H. M. '25. Psychological tests of musical talent. Rochester, N. Y.: Univ. of Rochester. Pp. 48.

—'28a. Seashore measures of musical talent. Psychol. Monog. 39, No. 178. 135–144.

—'28b. Measuring musical talent. Seashore tests as administrative aids. Personn. J. 7, 286–292.

Starr, A. S. '23. The diagnostic value of the audito-vocal digit memory span. Psychol. Clin. 15, 61–84.

Steacy, F. W. '19. The interrelations of mental abilities. Teach. Coll. Contrib. to Educ. No. 95.

Stenquist, J. L. '22. Stenquist mechanical aptitude tests. World Book Co.

—'23. Measurements of mechanical ability. Teach. Coll. Contrib. to Educ. No. 130. Pp. 101.

Strong, Jr., E. K., '22. A brief introduction to psychology for teachers. Baltimore: Warwick and York. Pp. 491.

Stutsman, R. '26. Performance tests for children of pre-school age. Gen. Psychol. Monog. No. 1, 3–67.

Sunne, D. '24. Comparison of white and Negro children in verbal and non-verbal tests. School and Soc. 19, 469–472.

Sylvester R. H. '13. The formboard test. Psychol. Monog. 14, Pp. 56.

Terman, L. M. '16. The measurement of intelligence: an explanation of, and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale. Boston: Houghton-Mifflin. Pp. 362.

Thorndike, E. L. '19. A standardized group examination of intelligence independent of language. J. Appl. Psychol. 3, 13–32.

Toops, H. A. '23. (with others). Tests for vocational guidance of children thirteen to sixteen. Teach. Coll. Contrib. Educ. No. 136, Pp. 159.

Town, C. H. '13. A method of measureing the intelligence of young children. (Trans. from Binet-Simon). Chicago: Chicago Med. Book Co.

—'21. An analytic study of five and six-year-old children. Univ. of Iowa. Stud. No. 48.

Trabue, M. R. and Stockbridge, F. B. '21. Measure your mind. Garden City: Doubleday Page. Pp. 349.

Turner, E. M. and Betts, G. H. '24. Laboratory studies in educational psychology. New York: Appleton. Pp. 218.

Wallin, J. E. W. '12. Experimental oral orthogenics. J. Phil. Psychol. etc. 9, 290–298.

—'12. Experimental studies of mental defectives. Baltimore, Md.: Warwick and York. Pp. 147.

—'16. Age norms of psycho-motor capacity. J. Educ. Psychol. 7, 17–25.

—'18. The peg form boards. Psychol. Clin. 12, 40–53.

—'21. Norms for the Seguin formboard based on the average of three trials. J. Delinq. 6, 381–386.

—'27. Clinical and abnormal psychology. Boston: Houghton-Mifflin. Pp. 649.

—'23. The measurement of traits in normal and epileptic school children. Miami. Univ. Bull. Pp. 175.

Weidensall, J. '16. The mentality of the criminal woman. Educ. Monog. Pp. 332.

Wells, F. L. '19. Psychologic performance in cancellation and direction tests. Psychol. Rev. 26, 366–371.

—'27. Mental tests in clinical practice. Yonkers New York: World Book. Pp. 315.

Wells, G. R. '17. Some experiments in motor reproduction of visually perceived forms. Psychol. Rev. 24, 322–327.

—'18. An apparatus for the mirror-drawing test. J. Educ. Psychol. 9, 99–101.

Wembridge, H. A. '23. Experiment and statistics in the selection of employees. J. of Amer. Statis. Ass'n. 8, 600–606.

Whipple, G. M. '10 and '14. Manual of mental and physical tests. Baltimore, Md.: Warwick and York. (Rev. Ed.) '14. Pp. 354.

—'15. Manual of mental and physical tests. Part II, Pp. 336.

Whitley. M. T. '11. An empirical study of certain tests for individual differences. Arch. Psychol. No. 19. Pp. 146.

Whitman, E. C. '25. A brief test series for manual dexterity. Educ. Psychol. 16, 118–123.

Wissler, C. '01. The correlation of mental and physical tests. Psychol. Rev. 3, 1–63.

Witmer, L. '11. Courses in psychology at the summer school of the University of Pensylvania. Psychol. Clin. 4, 245–273.

Woodworth, R. S. and Wells, F. L. '11. Association tests. Psychol. Rev. Monog. Suppl. 13, No. 57, Pp. 85.

Woolley, H. T. '15. A new scale of mental and physical measurements for adolescents and some of its uses. J. Educ. Psychol. 66, 521–550.

—'26. An experimental study of children at work and in school between ages of fourteen and eighteen years. New York: Macmillan. Pp. 762.

— and Cleveland, E. '23. Personality studies of three, four, and five-year-old children. J. Exper. Psychol. 6, 58–69.

— and Fischer, C. R. '14. Mental and physical measurements of working children. Psychol. Monog. 18, No. 77. Pp. 247.

Worthington, M. R. '26a. A study of some commonly used performance tests. J. Appl. Psychol. 10, 216–227.

—'26b. Performance test scores of behaviour and non-behaviour children. Welfare Mag. 17, No. 10. 97–103.

Yepsen, L. N. '29. The reliability of the Goodenough drawing test with feeble-minded subjects. J. Educ. Psychol. 20, 448–451.

Yerkes, R. M. '16. The mental life of monkeys and apes. A study of ideational behaviour. Behav. Monog. 3, Pp. 145.

—'21 (Ed.) Psychological examining in the United States Army. Mem. Nat. Acad. Sci. 15. Pp. 890.

—'21. A new method of studying the ideational behaviour of mentally defective and deranged as compared with normal individuals. J. Comp. Psychol. 1, 369–394.

—, Bridges, J. W. and Hardwick, R. S. '15. A point scale for measuring mental ability. Baltimore, Md.: Warwick and York. Pp. 218.

— Foster, J. C. '23. A point scale for measuring mental ability. 1923 revision. Baltimore, Md.: Warwick and York. Pp. 219.

— and Rossy, C. S. '17. A point scale for the measurement of intelligence in adolescent and adult individuals. Boston: Med. and Surg. J. 176, 564–573.

Yoakum, C. S. and Yerkes, R. M. '20. Army mental tests. New York: Holt. Pp. 303.

Young, H. H. '16. The Witmer formboard. Psychol. Clin. 10, 93–111.

—'16b. Physical and mental factors involved in the formboard test. Psychol. Clin. 10, 149–167.

—'22. Slot maze A. Psychol. Clin. 14, 73–82.

— and M. H. '23. The Witmer formboard first trial records. Psychol. Clin. 15, 85–91.

Young, K. '22. Mental differences in certain immigrant groups. Univ. Ore. Publ. Vol. I., No. 11, Pp. 103.

PART FOUR

SUMMARY, RECOMMENDATIONS AND CONCLUSIONS

Part Four

SUMMARY, RECOMMENDATIONS AND CONCLUSIONS

With Special Emphasis on the Results of the Discussion on Racial Comparisons

Summary

Two things have been attempted in this book: First, the various ways in which mental tests are used in experiments dealing with heredity and environment have been described, and second, a brief, analytical survey of non-verbal tests has been offered. This last might be considered as a first step toward a complete inventory of mental measurements, aiming to locate the best tests for use in eugenic research, which has been defined as "the study of the part played by human affairs in heredity".

PART ONE

Part One gave a historical sketch of the mental testing movement, and a brief description of modern conceptions of intelligence. Little can be concluded from these chapters, except that the field of mental testing is in a state of flux, and that intelligence theory has, until recently, been based on teleological considerations, rather than upon experiment.

PART TWO

Part Two discussed the general topic of mental "nature and nurture". Thus, in the first chapter were summarized experiments dealing with the effect on mental ability of different surroundings—especially the American home and school. The conflicting results from these studies (which were called "the analysis of environment") indicate that there is urgent need for more of their kind. They will ultimately furnish the best means of achieving tests of "innate ability", for when more is known about the effects of such environmental factors as schooling and language, achievement tests can be used to test the innate differences between particular groups. This for the reason that if a capacity can be inferred from an acquired accomplishment, a test of that accomplishment is a good test of that capacity.

The second chapter of Part Two described the various methods of investigating mental inheritance by means of mental tests. These were arbitrarily classified as "Retests", "Family Resemblances" and "Group Comparisons". The first was shown to be a crucial technique in the study of the permanence of traits and learning. The study of family resemblances seemed to furnish evidence in favor of heredity as against environment as a determining factor in human ability. This is hardly to be wondered at, since it is on blood relationships, after all, that evidence for heredity is ultimately based. The third technique—the comparison of groups—was dealt with by selecting racial comparisons as a specific example of inter-group testing.

The study of racial differences was given seemingly disproportionate treatment because it is a field in which mental tests seeking to measure innate ability have been most widely applied. Mental differences between races (in relation to our particular standards) have long been recognized;[80] and the interest of the eugenist is to discover how much they are due to the conditions peculiar to each group, and how much to biological causes. Blackwood ('27) says:

Racial Comparison

> "The difficulties of constructing tests adequate to compare the various peoples of the world seem almost insurmountable, including, as they do, the gigantic tasks of estimating the relative importance of the two great formative forces—heredity and environment—and of finding a means of separating the effects of one from the effects of the other." p. 14.[81]

Before this can be done, we have seen that pure races must be isolated, and of each race reliable samplings should be obtainable. Differences between groups in such factors as temperament, material culture, social conditions and language must next be taken account of, either by elimination or experimental control. The great majority of the tests used in comparing races are stamped with disapproval, because either their content, or the manner of their presentation is felt to be unfair to people of foreign background. The eugenist, the anthropologist, and the maker of immigration laws have all turned to the psychologist for help in measuring the mental elements in racial constitution. So far, psychology has offered only two scales intended primarily for inter-racial testing. While both of these present interesting possibilities, one has never been standardized, and the other must undergo radical improvements.

If a testing method were invented which could be applied as fairly to the rest of the world as to ourselves, real progress could be made in comparing races. Some suggestions for such a "universal" test were made in response to special questions for this report, and the most promising of these proposed that *learning tests* of one kind or another be used. On the whole, however, a wide-spread opposition to the idea of a universal yardstick was disclosed, and this seemed due to the fact that measurement in any sphere depends jointly on the existence of a measuring unit and on the homogeneity of the phenomena to be measured. (See Brown and Thompson '21). Even if a universal unit could be obtained, the world's population would never possess such homogeneity as mental testing demands, because the human mind is plastic, and is consequently moulded by the widely varying influences to which it is exposed. Because a "common" environment is so rare, and human groups are so unstable, even physical measurements cannot be standardized.[82] How then will it be possible to have universal norms for *intelligence?*

The question of universal testing seems to resolve itself to this: Just as all men can be measured physically, so can they be tested mentally, *within limits*. If we grant that sense discriminations are common to man the world over, and assume, further, that certain more complex experiences are also universal, then measurements of elementary reactions and possibly of some complex processes, if constructed from universal material, might furnish comparable results from widely differing civilizations. Of course, the argument that a test device depends on the culture which created it still holds good, and

Universal Testing

theoretically defeats the concept of a truly universal tool (See p. 47.) True universality of the higher mental functions is unthinkable, because it implies the disappearance of cultural diversities. When the Brotherhood of Man is attained, perhaps "psychologists" from all over the world will meet and build a test—or else not build one, if the Occidental concept of testing meets with Oriental disfavor! For the present we can say that as long as each nation is composed of several races, and differs from every other nation in geography and culture, these differences will have to be carefully studied and weighed by those who wish to compare races, and hence universal testing is circumscribed. This does not mean, however, that race has no importance for culture. Those who claim that efforts to gauge this importance are of no avail—that racial comparisons are odious—can be answered by saying that approximations toward universal, or at least international, tests are possible, and that they might prove to be useful tools for eugenics and anthropology, not to mention Congress. As Hankins ('26) puts it, the nervous systems of the human breed cannot have remained uniform throughout the course of evolution, and it is reasonable to suppose that differences in particular traits of intelligence and temperament do "affect political organization and history".

The list of universal items is easily exhausted and familiarity with actual testing technique can be made universal only by training. It is clear that these two facts decidedly limit the usefulness of a universal test, and that therefore a modified technique must be developed for measuring group differences, especially in the higher mental functions. The basic principle that tests are most efficient when best adapted to the groups on which they are to be used suggests that the ideal way to compare two groups would be to build a scale whose material would sample the environments of both groups, but through which would run a single test technique, so that the performance of each group on its own section of the scale would be comparable with that of the other. The problem of equating the difficulty values on such a scale however, renders this method out of the question at present. What does seem possible, is a method which has been termed "intragroup scaling". A list of items could be drawn up for two groups —A and B— omitting strictly localized material. These could be scaled on both groups, with the result that difficulty values of the various items would shift for each group. Two scales would result, similar in content and technique, but having different validities with respect to groups A and B. Both scales could then be used on both groups, and a real basis for comparison would be reached by comparing the two pairs of scores. If Group A did less well on Group B's scale in relation to its own, than Group B did on Group A's scale in relation to *its* own, then Group A would be the inferior one.

In the end, the cultural or racial origins of group differences can only be determined by completing five processes: (1) properly weighing biological influences such as natural and sexual selection: (2) determining at what age the final average level of mental growth is reached in each group; (3) carefully considering all external factors such as geography, the size of groups, migration, chance inventions and "culture" in the widest sense; (4) making sure that the samplings tested are reliable; and, last but not least (5) either (a) testing infants, or (b) developing "universal" tests, or rather tests which eliminate material not common to the groups being compared, or else (c) working out a system of "intra-racial scaling" which, as has been explained, would be both suited to the groups and valuable for comparisons. Even when all this is taken care of, however, comparisons will always have to be made in terms of *particular* characteristics. If race is to be defined as "a complex of traits inherited together within a limited range of variation" (Hankins '26), the average frequency of occurrence and the range of variability of a specific quality will be the basis for deciding on the superiority of one group over another.

PART THREE

Having seen in Part Two that tests which distinguish between learned and unlearned behavior are non-existent, and that therefore workers who wish to disentangle nature and nurture, must resort to the comparison of groups, it remained for Part Three to present the tests selected for the survey—tests more or less free from one of the major handicaps of group comparisons, namely, language. If it can be proved that the experience presupposed by non-language tests is more general than the training on which verbal tests are based—in other words, if such things as manipulative skill are less specialized than linguistic ability, then non-verbal tests can be profitably used as "tools of classification" within similar language groups. This kind of measurement, if done within a cultural milieu different from our own, would require the adaptation of the testing technique and material to the group in question. On the other hand, if non-verbal tests are experimentally proved to be merely tests of a specialized ability, and not tests of intelligence, they will continue to be useful for one of the purposes for which they originally were built, namely, the testing of subjects unfamiliar with the mother tongue of our country. They might also be used as "tools of comparison", to liken the "special performance ability" of peoples with different languages, provided said ability was found to be worth something, and also provided none of the tests contained strictly local material.

The first chapter of Part Three was devoted to this question of the validity of non-verbal tests as tests of intelligence. It was found that many authorities thought performance tests could not detect differences at the upper levels of intelligence—could not test abstract thinking—and that their results did not correlate highly enough with the results of verbal tests to be significant. On the other hand, when actual manual responses were examined, some were discovered to be far more abstract than others, so that the possibility arose that performance tests might not be limited to simple reactions. It was also found that the nearer the method of construction and validation of a performance scale approached the statistical requirements of a good verbal scale, the more defensible the former became as a test of intelligence. This would indicate that verbal tests do not monopolize "intelligence-testing-power", and that it may be possible to evolve non-verbal tests which would tap those faculties generally lumped together under the term "general intelligence". The most searching experimentation is needed, however, before this can be done.

Chapter II of Part Three briefly summarized the statistical standards to which test interpretation must conform. It was shown that judgments of individual performances on mental tests must be based on a knowledge of how the mass of comparable individuals performed on the same test, and that group comparisons must proceed along established statistical lines. Nothing need be said in this summary about the actual compilation of tests except to repeat the hope that it will prove useful to workers in many fields of measurement.

Recommendations

The second chapter of Part Three is intended to serve the research worker who desires to use mental tests for particular purposes. The science of eugenics proposes to undertake certain programs of research and should be in a position to ask the science of psychology for valid tests which could be adapted to this purpose. Before mental tests existed at all, sociologists expressed a practical need for measurements of individual differences, and, to meet this need, psychologists applied their theoretical knowledge and created tests for the insane, the feebleminded, the deaf and the delinquent. In the same way, the demands of anthropology resulted in the construction of a special scale intended primarily for testing primitive peoples. The above two instances are given in order to prove that it is possible for specialized sciences to co-operate, without either one encroaching illegitimately upon the other's territory. The most serious accusation against applied eugenics is that it has done this very thing. Herskovits ('29) referring to the "geneticists' and eugenists' proposals to control reproduction of humans so as to establish a more efficient social type" says:

> "The unity of the problem of human behavior is not to be thwarted by the erection of academic barriers, and we therefore have the spectacle of the biologist applying biological methods to the problems which lie entirely in the field of man's social activities, while the students of human society grope more or less blindly through the mazes of genetics. Biologists have sinned most grievously in attempting to cope with questions the solution of which lies outside the realm of their scientific equipment."

Realizing that genetics and eugenics cannot take up psychology's burden, and create tests, but at the same time realizing that the study of "human affairs and inheritance" necessitates entering into social fields—not for propaganda, but for controlled scientific experiment—it would seem necessary to establish a working partnership between the sciences involved. (See Young '24, pp. 47-48.) If eugenics is to ask other sciences for the facts it needs, and obtain them, there must exist central clearing houses of information in each science[83] analogous to such bodies as the Engineering Societies and the National Committee of Mental Hygiene. As regards mental tests, the first step in this direction would be to hold a conference of the leaders in this field, who would set up some sort of machinery whereby all existing data on testing methods and results would be available and kept up to date under expert psychological supervision. This would form a kind of library, and its yearly catalogue would constitute a sort of "Whipple's Manual". The modernizing of that famous mine of information is strongly advocated by Dr. W. V. Bingham,[11] who also suggests some central committee of experts who could be brought together from time to time to review the total situation, appraise the newer developments, and plan for constructive research leading to "invention, calibration, standardization and validation of needed mental tests." Dr. Whipple himself, in a letter to the writers, refers to the "really startling advance made in 1917 by bringing together the group which constructed the Army Alpha", and goes on to say that "a conference of perhaps a couple of days' duration, held annually for ten or a dozen persons immediately concerned in the practical use of intelligence tests. . . . would be of extraordinary value". In 1921 Colvin urged the establishment of a "committee of skilled psychologists to select the elements most valuable in the tests now existing, add others that are lacking, and carefully standardize the complete test".

The need for co-operation between the technical laboratory and the clinical field has often been voiced.[84] It is expressed thus, in reference to character measurement, by G. B. Watson ('26): "Cannot those whose interest has been the reliable observation of behavior come into closer relationship with those whose interest has been its meaningful interpretation?" That this can best be effected by real co-ordination in the field of mental tests has been the opinion of many writers and practical workers. As far back as 1916, Wallin proposed "a station of psychological tests and standards", whose yearly budget would approximate fifty thousand dollars. Haggerty ('21) likewise proposes "some scientific organization enjoying public confidence" in order that agreement could be reached on the "essential criteria which satisfactory tests should meet". Professor T. L. Kelley[85] expresses himself as favorable to the idea of a central bureau, but points out that it would be "inefficient unless heavily financed". Of like mind is Dr. Goodenough[85] who remarks that a central bureau would be useful only if "run by someone capable of evaluating the material, for a purely mechanical bureau would probably dispense as much misinformation as reliable data". According to Dr. Brigham,[85] such a bureau would be called "A Bureau of Prematurely Standardized Tests". A similar note of warning is sounded by Dr. L. D. Anderson, who writes:[85] "If such a bureau would deal in terms of reliability and validity of tests . . . . I would be in favor of it, but in general the presence of test material in a central bureau carries the implication that the test is worth something." Dr. Kent,[85] who advocates a central test library where all test scores and records would be filed, says in regard to test standardization: "Much depends upon what tests are to be standardized and by whom the work is to be done". Other psychologists who work with tests have, in answering the test data sheets sent out for this report, registered interest in any project which might co-ordinate the work of human mental measurement, viz: Drs. E. L. Thorndike, F. L. Wells, P. Blanchard, G. Hildreth, F. N. Freeman, G. O. Ferguson, Jr., H. H. Goddard, S. C. Kohs, F. N. Maxfield, P. C. Squires, E. S. Jones and J. P. Herring.

Concrete examples indicating the need for such a bureau are the following:

1. One of the authors of the "Manual of Individual Tests" (Bronner, Healy, Lowe, and Shimberg) has remarked on the difficulties encountered in gathering information on individual performance tests. It took about two years of research to assemble 126 tests, and snarls were encountered in such matters as: (a) locating the widely scattered and often unnamed tests; (b) ascertaining which variations and modifications the authors had in mind, and to which of these variations the different norms applied; (c) obtaining directions as to how the tests should be given.

(The present writers, collecting tests of much the same class, could elaborate these difficulties ten-fold, and herewith feelingly register a complaint against the present lack of system in nomenclature and standardization of tests.)

2. A German, discussing the development of a universally usable psychometric system, cites as the chief difficulty the fact that test results are rendered incomparable by the large number of measuring units (in Germany alone). Before this condition can be remedied, a collection of all tests in present use would have to be made. (Szondi '29).

3. A psychologist, Dr. R. Willoughby, doing an individual piece of genetic research, remarked to one of the writers that he had no means of connecting himself with

the unpublished work being done currently. He stressed the need for a dependable co-ordinating agency which could give him both a survey of all methods dealing with his problem which have already been used or proposed, and also giving recommendations with respect to his special needs.

4. An important manufacturer of test material is still listing in his yearly catalogue many tests which are out of date, and which have never been properly standardized. The tendency of test constructors to turn out inadequate material, of supply houses to urge its use, and of test users uncritically to accept it, can only be checked by some central bureau of information.

5. A recent dissertation (Shimberg '29) deals with the necessity for analyzing test results in the light of specific group surroundings instead of on the norms yielded by the group on which the test was standardized (probably hundreds of miles away, with entirely different conditions of life). This points again to the need for a central agency which would have records of the various factors entering into the standardization of a test. (These would include a "consideration of the group tested, the variation of the norm with the addition of each group of results, and the type of standardization required.")[86]

As far as the present writers know, only one extensive attempt has been made to collect and classify tests.[87] This takes into account only pencil and paper tests which have already been printed and standardized; it omits performance test material (at most, filing only the manufacturer's catalogue), experimental work, and tests in process of standardization. A bureau such as is here proposed would include, besides actual test material, records of all investigations which make use of mental tests and without which tests mean almost nothing. *It would not be a Bureau of Standards in any sense of the word, for it would be based on the supposition that the most important work at present is experimentation with the actual tests themselves, so that the theoretical implications underlying them shall be clarified.* If this supposition gained wide credence, piled up standardizations of inferior tests would be shown up as so much wasted effort. At the risk of seeming premature, the writers have drawn up a prospectus for a central Bureau of Information on Mental Measurements. It is to be found in Appendix II, together with lists of experimental and unworked material, and tentative suggestions and needs.[88]

If such a bureau were established it would serve psychologists, clinical workers, teachers and sociologists just as much as it would serve the eugenists, if not more. For this reason, it should be under the auspices of authorities in these different fields. This is in accordance with the fact that eugenics stands in the unique position of having to draw its experimental material from biology on one hand and human institutions on the other. The "science of eugenics" is here thought of as distinct from "applied eugenics" on the assumption that the latter depends on the former. As Campbell ('29) puts it:

> "Eugenics, both in the field of investigations and in the field of application, is largely dependent upon its liaisons . . . In addition to the biological affiliations of eugenics, it is in the best interest of eugenical investigation to maintain liaisons with all of the social sciences."

Just as the embryologist must know about the temperature of his laboratory, so must the eugenist make an effort to learn about all the possible factors influencing heritable human traits. Although he does not go so far as the behaviorists, and assert that there is no real unlearned response except for the first movement of the fertilized egg,[89] he recognizes that the strength of different traits (after early infancy) is due both to "heriditary propulsion" and to the individual's experience, and that hereditary influence is thus only displayed when stimulated by the environment. So, although his primary interest is the discovery of what is hereditary and what is not, the eugenist seeks to learn about social stimuli as well as about biological mechanisms.

He cannot pursue his work however, without valid tools and standards of measurement, and in the present disorganized state of mental tests, those tools and standards are extremely difficult to obtain. It is believed that he would be greatly aided in his task (when it has to do with mental traits) if some unifying agency, such as this Bureau, were in existence. Francis Bacon describes the purpose of the research foundation which he made the center of his fabled paradise in the *New Atlantis,* as being "the knowledge of causes and secret motions of things; and the enlarging of the bounds of human empire, to the effecting of all things possible." If our imagined Bureau performed its work comprehensively, it might approach the discovery of the Reason which lies behind the great fact of individual differences, and causes the human mind to be

> . . . . in some maybe
> stagnant and poor, in some activ and rich, in each
> a given unique quantum of personality,
> a loan of so-much (as 'tis writ *to one he gave*
> *five talents, to another two and to another one).*
> Bridges '29, (Part II, lines 717–721)

Final Conclusions

With the above question no one living is qualified to deal. All this book has sought to show is that intelligence tests are still handicapped as measurements of innate mental differences by the fact that the effects of training, deeply influencing test performance, are not yet ascertainable. The fact has been stressed that

> ". . . . caution should be exercised . . . . in employing tests to study the nature-nurture question . . . . (because) careful research workers assert that the mental tests are measures of mentality *only in so far as mentality is reflected in educational achievement.*" (Witty and Lehman '30. Italics ours. See also Burt '21, p. 175).

No "hereditarian" insists that tests measure a fixed innate power, while even the most extreme "environmentalists" admit that biology places limits on every performance. Psychologists are unwilling to admit that any test is free from individual experience, and the general opinion is that it is impossible to construct such a test at present.[90] The term "test of innate ability" must be taken as meaning a test which, by taking account in some way of environmental differences in specific situations, can discriminate between persons of different capacities.[91] There is no qualitative difference between tests of capacity and achievement; the former merely contain less "deliberatly" learned material than the latter. The relationship which must be made clear is that existing between two general pairs of factors, of which one in each pair always operates when an individual is tested. In our civilization, these are, generally speaking, a "cultured" home and a good school as against a limited home and a poor school, and native cleverness as against native stupidity.

The influence of nature and of nurture can never be disentangled within an individual, due to the "inescapable abstractness" of truly native intelligence. (Valentine '29). Or, as Cattell, puts it: "What a man can do is determined by his native equipment, and what he does do, by the circumstances of his life." ('24) If, however, we ignore the individual, and think in terms of groups, it may be possible to control either environmental or hereditary factors so that the effects of one are distinguishable from the effects of the other in specific instances. But even after the scores of groups differing in race, geography, social class, schooling and vocation have been compared, the differences between them can be interpreted alike by the "hereditarians" and the "environmentalists". The line of cleavage between these two schools is therefore gradually fading out; protagonists in each camp are realizing that "native" and "acquired" are merely two adjectives serving to describe two extremes of a continuous scale. The issue between these two extremes can never be resolved for all times and places. Jennings ('30) says: "Environmental diversities as certainly affect the characteristics of organisms as do genetic diversities. . . . To determine to which category a given diversity belongs requires a knowledge of the pertinent concrete facts". It is futile to generalize about the heredity-environment question. It must be dealt with specifically, with the object of discovering "how two variables can best be controlled to obtain a desired result". (Kirkpatrick '26) In the words of F. S. Freeman ('29) this issue must cease to be a controversy, and must become, rather, a "problem of growth or development, and of its direction or control." He quotes no less an authority than Plato: "In the case of all seeds, and everything that grows, we know that whatever fails to find its appropriate nourishment, season and soil, will lack its proper virtues." (Republic, 481)

Granting the basic importance of heredity, the real question becomes this: If environment can raise or lower the I.Q. by twenty points, is not such a power of vital consequence to each nation?[92] Is experience — i.e. education — still "the mother of wisdom", so that, in the words of John Dewey ('22) "we shall never get any light upon what are the limits to intelligence set by innate qualities till we have immensely modified our scheme of . . . giving experience"? Galton ('83), defending his attempt to appraise the relative importance of heredity, says: "We know that the bulk of the respective provinces of nature and nurture are totally different, although the frontier between them may be uncertain." (p. 182) If man's development is ultimately to be controlled by man, *it will have to be accomplished in this borderline area.* At present, the most we can say is that

> "Nature is more important than nurture in explaining individual differences in acquired intelligence when the nurture has been similar for the group concerned. On the other hand, it would be equally true that nurture would be more important than nature in explaining individual differences if the native equipment of a group were substantially the same, and the environment markedly different." (Colvin '22, p. 181)

Note on the Twenty Seventh Yearbook of the National Society for the Study of Education.

Taken chiefly from the Journal of Educational Psychology, 1928, Volume 19.

Since so many references have been made in the preceding section to the contents of this famous Yearbook, a few quotations summing up its contributions to the question of heredity and environment seem to be in order. In it is to be found "crucial evidence in favor of or against either nature or nurture as determiners of mental capacity or of school achievement" (Burks), in the form of 39 new research projects, many of which were embarked upon at the instigation of the Society's "Committee on the Possibilities and Limitations of Training". The articles are divided into those dealing with the influence of nature and nurture upon intelligence (Part I), and upon achievement (Part II). (However, as Gates points out, "the problems in both volumes are concerned with the influence of various factors upon achievement", since native ability can only be gauged through performance.) Baldwin regards the Yearbook on the whole as proving that "learning and maturation are dependent on each other", and that "maturation is influenced, accelerated, retarded and modified by environmental factors". Both Judd and Freeman regard the data as proving that environment is by no means a minor factor in intelligence test scores. Freeman finds that the child's performance on a test is influenced to a marked degree by the character of the home he has been reared in, and Judd concludes: "Individuals are what they are because of what they inherit from their germ plasms and from their social surroundings". Gates finds that the studies in the Yearbook show that "native endowment results in different rates of learning", while Whipple goes so far as to intimate that the evidence supports the "hereditarians". Finally, the editor, Terman, says: "Part I of the Yearbook shows that the I.Q. does count, is not easily influenced by environmental factors, and is relatively constant .... although not entirely beyond the influence of environment .... With respect to the whole nature nurture problem .... there is relatively little which one can set down as precisely established fact".

Witty and Lehman (See p. 10) criticize the Yearbook findings because they claim these are based on two uncertain assumptions: (1) That "mental tests measure general intelligence", and (2) that "mental traits are subject to the same laws of inheritance as are physical ones." Walter Lippman levels identical criticism at mental testers in general. (See *The New Republic*, 1922, Vol. 33, pp. 9–11, and the answer by Terman, pp. 116–120). For discussion of these two controversial questions, see pages 10ff and 31ff of the present volume.

BIBLIOGRAPHY

References cited in Part IV.

Summary, Recommendations and Conclusions.

Blackwood, B. '27. A Study of mental testing in relation to Anthropology. Ment. Meas. Monog., 4, Pp. 120.

Bridges, R. '29. The testament of beauty. Oxford: The Clarendon Press. Pp. 192.

Brigham, C. C. '30. Intelligence tests of immigrant groups. Psychol. Rev. 37, 158-165.

Brown, W. and Thomson, G. H. '21. The essentials of mental measurement. Cambridge: Cambridge Univ. Press (3rd Ed.) Pp. 216.

Bruner, F. G. '14. Racial differences. Psychol. Bull., 11, 384-386.

Burt, C. '21. Mental and scholastic tests. London: County Council. Pp. 432.

Campbell, C. G. '29. The liasons of Eugenics. Eug. News, 14, 33-37.

Cattell, J. Mck. '24. The interpretation of intelligence tests. Sci. Mo., 18, 508-516

Colvin, S. S. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol., 12, 136-139.

— '22. Principles underlying the construction and use of intelligence tests. 21st Yrbk. Nat'l. Soc. Stud. Educ. 11-45.

Darsie, M. L. '25. The mental capacity of American born Japanese children. Comp. Psychol. Monog., 3, Pp. 89.

Davenport, C. B. '29. Do races differ in mental capacity? Human Biol. 1, 70-89.

Dearborn, W. F. Shaw, E. A. and Lincoln E. A. '23. A series of performance tests of intelligence. Harvard Monog. Educ. 1, No. 4, Pp. 64.

Dewey, J. '22. Mediocrity and the individual. New Rep., 33, 35-37.

Dodd, S. C. '26 International group mental tests. Princeton, N. J.: Princeton Univ. Store, Agents. (Mimeographed). Pp. 101 plus appendix.

Freeman, F. N. '26. Mental tests. Boston: Houghton Mifflin Co. Pp. 503.

Freeman, F. S. '29. Intelligence tests and the nature-nurture controversy. Sch. & Soc., 31, 830-836.

Galton, F. '83. Inquiries into human faculty and its development. New York: MacMillan. Pp. 387.

Goodenough, F. L. '26. Racial differences in the intelligence of school children. J. Exp. Psychol. 9, 388-397.

Graham, J. L. '30. A quantitative comparison of rational responses of Negro and white college students. J. Soc. Psychol. 1, 97-121.

Graham, V. T. '26. The intelligence of Italian and Jewish children in the habit clinics of the Massachusetts Division of Mental Hygiene. J. Abn. & Soc. Psychol., 20, 371-376. See also Graham's studies on Chinese children (J. Comp. Psychol. 6) and on Negro children (Pub. Health Rep. 1926, 2757-2783.)

Haggerty, M. E. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol., 12, 212-216.

Hankins, F. H. 26. The racial basis of civilization. New York: A. A. Knopf. Pp. 384.

Herskovits, M. J. '29. Social selection and the formation of human types. Hum. Biol., 1, 250-262.

Holmes, S. J. '29. The interplay of heredity and environment. Child. Stud. 6, 139-141.

Holmes, S. J. '30. Nature versus nurture in the development of the mind. Sci. Mo., 31, 245-252.

Johnson, C. S. '30. The Negro in American civilization. New York: H. Holt.

Kirkpatrick, C. '26. Intelligence and immigration. Ment. Meas. Monog., 2. Pp. 127.

Klineberg, O. '28. An experimental study of speed and other factors in "racial" differences. Arch. of Psychol., 93, Pp. 111.

Kuo, Z. Y. '29. The net result of the anti-heredity movement in psychology. Psychol. Rev. 36, 181-199.

Lenoire, Z. '25. Racial differences in certain mental and educational abilities. Thesis, State Univ. of Iowa Library.

Mayo, M. J. '13. The mental capacity of the American Negro. Arch. of Psychol. 28.

Murdoch, K. '20. A study of race differences in New York City. Sch. & Soc., 11, 147-150.

Peterson, Jos. '28. Comparison of white and Negro children in the rational learning test. 27 Yrbk. Nat'l. Soc. Stud. Educ. Pt. I, 333-341.

—and Lanier, L. H. '29. Studies in the comparative abilities of whites and Negroes. Ment. Meas. Monog., 5. Pp. 156.

Porteus, S. D. and Babcock, M. E. 26. Temperament and race. Boston: Badger. Pp. 364.

Price, J. St. C. '29. The intelligence of Negro college freshmen. Sch. & Soc. 30, 749-754.

Rigg, M. '28. Some further data on the language handicap. J. Educ. Psychol., 19, 252-256.

Sandiford, P. and Kerr, R. '26. The intelligence of Chinese and Japanese children. J. Educ. Psychol., 17, 361-367.

Schwegler, R. A. and Winn, E. '20. Comparative study of intelligence of white and colored children. J. Educ. Res., 2, 838-848.

Seago, D. W. and Koldin, T. S. '25. A comparative study of the mental capacity of sixth grade Jewish and Italian children. Sch. & Soc., 22, 564-568.

Shimberg, M. '29. An investigation into the validity of norms with special reference to urban and rural groups. Arch. of Psychol. 104, Pp. 84.

Stormzand, M. J. '22. Intelligence tests and eugenics. J. Ap. Soc. 6, June–July.

Strachan, L. '26. Distribution of intelligence quotients of 22000 primary school children. J. Educ. Res. 14, 169-177.

Szondi, L. '29. Zur Psychometrie der Tests. Arch. f. d. ges. Psychol. 72, 43-114.

Thorndike, E. L. et al '26. The measurement of intelligence. New York; Columbia Univ. Publ. Pp. 616.

Valentine, P. F. '29. Beyond heredity and environment. Educ., 50, 65-74.

Viteles, M '28. The mental status of the Negro. Ann. Amer. Acad. Pol. & Soc. Sci., 140, 166-177.

Wallin, J. E. W. '16. Mental tests. A Symposium. J. Educ. Psychol. 7, 353-357.

Wallis, W. D. '26. Race and Culture. Sci. Mo. 23, 313-321.

Watson, G. B. Character tests of 1926. Voc. Guid. Mag. 5, No. 7.

Whipple, G. M. '22. Intelligence tests in colleges and universities. 21st Yrbk. Nat'l. Soc. Stud. Educ.

Witty, P. A. and Lehman, H. C. '30. The dogma and biology of human inheritance. Am. J. Soc., 35, 548-563.

Woodworth, R. S. '16. Comparative psychology of races. Psychol. Bull., 13, 388-396.

Woolley, H. T. '26. An experimental study of children at work and in school between ages of fourteen and eighteen years. New York: MacMillan. Pp. 762.

Yeung, K. T. '21. Intelligence of Chinese children in San Francisco. J. Appl. Psychol. 5, 267-274.

Yoder, D. '28. Present status of the question of racial differences. J. Educ. Psychol. 19, 463-470.

Young, K. '22. Mental differences in certain immigrant groups. Univ. of Oregon Publ., 1, No. 11. Also Sci. Mo. 15. 417-434.

—'24. The history of mental testing. Ped. Sem. 31, 1-48.

APPENDICES

APPENDIX I

FOOTNOTES IN TEXT

Introduction

(1) Page vii. Galton's classic definition of eugenics runs: "Eugenics is the study of agencies under social control which may improve or impair the racial qualities of future generations either physically or mentally."

Prof. Carr-Saunders points out that this is not a definition of a pure science, since it introduces the words "improve" and "impair." He says: "In order to decide whether any agency is improving or impairing racial qualities, it is necessary to introduce judgment of value, to decide what is good and what is bad. With good or bad a pure science has nothing to do; it is concerned only with what is." Applied eugenics, defined by Galton, is thus a separate thing from the science of eugenics as we know it now. The best definition for the latter at present is "The study of the part played by human affairs in heredity"; whereas "the study of the part played by heredity in human affairs" seems to be in the field of anthropology and sociology. The science of eugenics should concern itself with the study of the effect of social phenomena on the processes of biological evolution. This is a proper subject for scientific enquiry, which should be free from such assaults as are levied against applied eugenics.

(2) Page vii. Parts I, II, and IV were written by B. Schieffelin, with the assistance of G. Schwesinger, while Part III was compiled by G. Schwesinger with the assistance of B. Schieffelin.

PART I.

Chapter II, Theories of Intelligence.
Bibliography to be found on pages 16—18

(3) Page 10. See Stockton '22 and Boring '23 and Pyle '28. This opinion was also expressed by Binet (see page 5 of this volume) and by Drs. R. Pintner and F. Boas in personal interviews. Prof. Thorndike has stated that when an investigator is required to demonstrate just what abilities are measured by his scale of tests, there is recourse in one sure answer only. This consists in publishing a detailed description of the tests, together with a full account of the testing conditions, methods of scoring and statistical treatment of test results.—'24, pp. 219-252.

(4) Page 11. Longer descriptions of intelligence are to be found in "Intelligence and its measurement: A Symposium." J. Ed. Psych. Vol. XII. 1921. Pages 123-147; 195-216. See also Pintner '23, pp. 52-60 and Freeman '26, Chap. XVIII.

(5) Page 11. This definition is the opposite of the definition of feeblemindness given by the British Royal Commission: "A feebleminded person is one who is capable of earning a living under favorable circumstances, but is incapable, from mental defect existing from birth, or from an early age, (a) of competing on equal terms with his normal fellows; or (b) of managing himself or his affairs with ordinary prudence."

(6) Page 12. More recent contributions to the literature on intelligence theories are: Ko '27; Black '28; Carroll '28; Edwards '28; Claremont '28; Stern '28; Gavit '29; Lashley '29 and Laycock '29. (This last has to do largely with Spearman's theory.)

(7) Pages 12–13. Freeman '26 (p. 483) distinguishes between quantitative and qualitative accounts of intelligence by citing as an example of the latter, William James' description of the way a man seeks for the cause of a smoking lamp in order to fix it. This is not intended to show the "different degrees of intellectual capacity" but rather "the essential nature of intelligent behavior . . which characterizes intelligence wherever it manifests itself, whether in high or low degree."

(8) Page 13. This information was obtained at the Psychological Clinic. (Dr. M. Murphy, personal inerview). Dearborn et al '23, cite in this connection the example of Bradley's discovery of the reason for the aberration of star light. He was sailing up the Thames, and observed a change in the position of the vane on top of the mast. He inquired the reason of the sailors, and was told that the change was due to a change in the position of the boat, not to a shift of wind. Dearborn concludes: "that the sailors reasoned of wind and boat, and the scientist of light and stars, would not seem a material difference in judgment of the intelligences involved. But as a result of added information the scientist takes a second step—association by similarity. If, however, the mariner in turn through a new association by similarity applies the experience of the scientist to the directing of his boat, who will offer to decide as to the relative intelligence involved?"

(9) Page 13. Henmon '21 suggests that capacities should be given weight (i. e. importance in General Intelligence) in progortion to their usefulness, which he defines as "ability to attain success in general." This would naturally give more weight to the complex processes. In reference to classifying intelligence into verbal, concrete and social types, Freeman '26 says: "This rather widely current distinction lacks statistical support from the results of tests, and it is at least possible that the variations in abilities to deal with symbols, persons and things may be explained in another way."(p. 480).

(10) Page 13. See, for instance, Pear '22, Chap. XII; Pintner and Upshall '28 and Dwelshauvers '28. See also M. R. Trabue's communication, Ninth International Congress Psychol. 1929, in which he makes the point that "intellectual power to handle one type of material (e. g. politics) may not be superior to that required to deal successfully with other types (e. g. plants or animals).

(11) Page 14. Told to one of the writers in a personal interview.

(12) Page 15. Viz. Keller and Brigham. For recent articles of interest in relation to group factors, see Farmer '27; Dodd '28; Mackie '28; Brigham '28 and Schneck '29.

(13) Page 16. "Men who are plodders but careful and accurate thinkers"—Colvin. (Ref. Pintner '23, p. 279). Thurstone '21 says: "Intelligence tests do justice to the well-balanced bright mind, not to the more inhibited profound type of intellect", and again: "Volitional drive, and instinctive energy must be associated with analytic capacity.... Intelligence is only one of the elements in mentality, and it has been overworked because of being accessible to measurement." Porteus and Babcock '26 have devised the term "psychosynergic" to describe the "factors outside the cognitive" which give force to character by "balancing or cooperating mental energies." Pyle '28 emphasizes the fact that intelligence varies in the same individual, depending on circumstances and emotional sets. See Colvin '22; Bruce '22; Beauchamp and Webb '27; Mackaye '28; Meili '29; Oates '29 and the communication by P. A. Vernon at the 9th International Congress of Psychology.

PART II.

Introduction. Bibliography to be found on page 22.

(14) Page 21. The expressions in the text were adapted from Holmes '21, p. 19, and also from a lecture given at Columbia University by Prof. R. S. Woodworth. E. G. Conklin '23 has summed up the situation thus: "The development of an individuai is dependent upon the interaction of two sets of factors or causes, the intrinsic and the extrinsic. The former is represented by the organization of the germinal protoplasm and the latter by all other conditions... The great problem of development is the unravelling of these two factors, the assignment of its true value to each, and the ultimate control of development so far as this may be possible." See also Elderton '09; Pearson '10; Davenport '11; Thorndike '13; L. Darwin '13; Popenoe '15; Kellogg '23; Jennings '24; Child '24 (Chapters 1-4); Carr '25, (Chapters 6 and 7); Perrin and Klein '26; Ogden '26; Dunlap '27; H. L. Hollingworth '27; Gates '28; Terman '28; Whipple '28; Draper '29; and Lowie '29. See also pages 252–253 of this volume.

Chapter I. The Measurement of Environment.

Bibliography to be found on pages 26–30.

(15) Page 23. Gates '28, explaining why certain tests are called "intelligence tests" says: "When educative forces for a number of persons of the same age are absolutely identical, individual variations in ability appear which under these conditions, may be

attributed to differences in native constitutional factors." (p. 383) He then goes on to show that educative influences in a typical American community are near equality.

Pintner-Paterson '17 say: "The difficulty of finding anything that is not influenced by education in school is well recognized, and we are forced to take for granted the acquisition of such general abilities as reading and writing in children that grow up in the ordinary civilized community." Terman '16 shows how minor environmental differences do not affect the validity of tests. (p. 116) The writers have come across one case in which an individual who lacked home training, normal intercourse with other children and all but the most rudimentary education, was studied: Foster '19 reports a boy who tested normal on the Binet and Yerkes-Bridges scales, despite the above disadvantages.

(16) Page 23. Among the exponents of the former attitude (at least in the case of whites and Negroes) are Strong '13 and G. R. Wells '23. Bridges and Coler '17 and Fukuda '25 give the reader the same impression, but do not place environment and intelligence in causal sequence. Examples of studies tending toward the latter explanation are Weintrob '12; Pressey and Thomas '19; L. W. Pressey '20; Courtis '26 (a); Hirsch '28; Goodenough '26. No one of this latter group of studies, however, postulates a single principle of causation, and nearly all acknowledge the weakness of tests as measures of innate ability. See also Fryer '22, who actually lists occupational intelligence standards, and the Barr Scale of Occupational Intelligence in Terman's "Genetic Studies of Genius".

(17) Page 23. This point was brought up at a general discussion at the Ninth International Congress of Psychology, Conference on Race Differences. See Bickersteth and also the studies in England by Thompson, and Duff and Thomson on selection as against environment. (For annotated references to these and other studies on the relation between intelligence and occupational status, see Burks '28 (b), pp. 268-273. Especial attention is called to the study by Haggerty and Nash.)

(18) Page 24. See Yerkes '21 (Chap. X); Burt '21 (pp. 175-194); Willard '22; Gates and La Salle '23; Chapman '24; Bagley '25; Wechsler '26; Kelley '26; Hildreth '28; Chen '28; Goodenough '28; Davis '28 and Burks '28 (b), pp. 296-395.

(19) Page 24. See Rugg and Colloton '21; Teagarden '22; Dvorak '24; Hurlock '25; Nettles '26; Hildreth '26; Gray and Marsden '26; Broom '27; Cornell '27; Randall '27; Slocombe '27; Matthew and Luckey '28; and Burks '28 (b), 319–325. See also the work of Minogue and Otis on the constancy of the I.Q. for the feebleminded, and of Duff and Terman for the superior. T. G. Foran has reviewed work done on the constancy of the I.Q. up till 1929 (Cath. Univ. Amer., Educ. Res. Bull. 4, No. 9).

(20) Page 24. Environment taken in its widest sense includes such diverse factors as climate, physique, maturity and emotional traits. See Burks '28 (b), pp. 305–319. See also Kempf and Collins '29; Kiefer '29; and also the studies in the 27th Yearbook by Gesell, Hoefer and Hardy, Stone and Doe-Kuhlman, etc. An extensive retesting program is now being undertaken by the Psycho-educational Clinic, Cambridge, Mass. It is known as the Harvard Growth Study, and involves the measurement of a great number of children with intelligence and achievement tests over a period of ten years. See Dearborn '24 and Cattell and Gaudet '30.

Dr. Francis Maxfield, in a personal interview, suggested a practical scheme whereby the effect of different surroundings on the same people after a lapse of time could be gauged. If a group of children found to be homogeneous at the age of five or six, were retested ten years later, when some would be in the country, others in the city, something could be learned as to the supposedly different effects of urban and rural surroundings.

(21) Page 24. For annotated references to studies on intelligence and social environment, see Burks '28 (b), pp. 273–282. See also Book '22, Bagley '23; Stern '24; Colvin and MacPhail '24; Woolley '25; Macdonald '25; L. S. Hollingsworth '26; Jones and Carr–Saunders '26; Decroly '26; Blanchard and Paynter '27; Gesell and Lord '27; Aldrich '27; Furfey '28; Goodenough '28 (b); Goodenough and Shapiro '28; Hetzer and Reindorf '28; Stroud 28; Chauncey '29; Sirkin '29; Van Dael '29 and Probst (unpub.). For studies of the relation between I.Q. and size of family, and position in family, see Bradford '25; Chapman and Wiggins '25; Sutherland and Thomson '26; Arthur '26 and Lentz '27. C. R. McRae (Australia) has made an unpublished study on "some effects of social and educational opportunities upon mental tests."

(22) Page 26. Taken from Burdick '28. According to Goodenough the real question is "To what extent does the increased ability to deal with the type of task included in these tests which seems to be brought about through more effective environmental opportunity, correlate with increase in the ability 'to think in abstract terms', to

'make correct responses', 'to adjust oneself to his environment', or to 'act better in novel situations' outside the test situation?" (Conference of the Child Development Committee of the National Research Council). See also Pressey '20. p. 92, Courtis '26 (b) and Fukuda '25, p. 130. A. M. Conklin '30 also stresses the need for finer instruments with which to analyze family situations.

Chapter II. Studies of Mental Inheritance.
Bibliography to be found on pages 52 – 58.

(23) Page 31. Taken from Baldwin '28 and Burks and Kelley '28, p. 30 respectively; Jennings '25 seems to prove that mental unit characters are not inherited in man, and that this fits in with the supposition that human characteristics are molded by environment within wide biological limits. Goddard '29 holds that mental aptitudes rather than traits are inherited: "There is no spiritual entity in the germ plasm." H. E. Jones '28 says: "While not in agreement with a naive unit-inheritance theory of intelligence, our results are undoubtedly amenable to statement in terms of multiple Mendelian factors, additive complexes of genes, such as have been assumed to function in the inheritance of quantitative traits, in plants and animals." He refers here to Pearson's article "On a mathematical theory of determinantal inheritance."— (Biometrika, Vol. 6, pp. 80–93) See also the discovery by two Germans (Professors Haecker and Ziehen) to the effect that musical talent is transmitted according to "hereditary laws", (Ref. Science, n.s. 57; supp. 12, June 8, 1923) and Fisher's study of family correlation of mental traits ('18).

(24) Page 31. For example, the segregation of mental traits can be studied in a case where feeblemindedness is latent in two parents and appears as a mental defect in one child, while his brother is normal. (See Goddard '13) T. H. Morgan '29 says: "For the present purposes, the occurrence of such sporadic physical defects that are inheritable furnishes at least a basis for opinion as to the likelihood of similar conditions affecting the psychical make-up of human individuals", but in another place he stresses the hazards of such a supposition by pointing out that man is uniquely affected by his surroundings because of his enormous capacity to learn, and consequently "no definite conclusions can be reached until we know more about the relative importance of heredity and environment in each specific case."

Burks and Kelley '28, p. 31, speaking of a group of studies of physical heredity say that "they supply the cues for investigations of mental heredity." But later Kelley '30 denies that the methods of the geneticist can be used to study mental heredity.

Jennings '30 says: "It is positively known that gene diversities bring about great differences in mentality," and over two decades ago, Pearson '03 concluded: "We are forced, I think literally forced to the conclusion that the physical and psychical characters in man are inherited within broad lines in the same manner and with the same intensity." See also Pearson '01, Woodrow '19 and Kirkpatrick '26. (In the latter is to be found a review of studies dealing with the analogy between mental and physical traits. See also the reference to the study of Witty and Lehman, on page 10 of his volume.)

(25) Page 31. "Indeed a case could well be made out for the thesis that the theoretical objections sometimes brought against mental measurement really hold in the last resort against all measurement, and prove too much; and that the real difference between mental and physical measurement is simply that mental phenomena, being practically more difficult to handle, force on our notice the epistemological difficulties inherent in all measurement, whereas in physical measurement, familiarity has bred contempt." Brown and Thomson '21, p. 12.

(26) Page 32. In another part of the 27th Yearbook (Pt. II, p. 282) Burks says: "Until we can have controlled experiments upon children of various ages transplanted at infancy into uniform environments, precise knowledge regarding native mental differences may be impossible to secure." Of interest in this connection is the work of Feingold '24, who tested the first generation of a group of European immigrants and found that the gap between their parents and native Americans (as ascertained from Army test results) was more than half bridged. He concludes that most of the original gap was due to environment. See also the reference lists on the effect of schooling and coaching on test results, and the constancy of the I.Q. (Burks '28b) Kelley's definition of "nature" is

"a trait making for individual differences which does not change with age." (See reference on page 24 of this volume.)

(27) Page 32. This is taken from Dr. F. Goodenough's paper read at the 1929 Conference of the Child Development Committee of the National Research Council. One study of the effect of frequent retests at short intervals on a variety of tasks was made at the Boston Psychopathic Hospital during the winter of 1928–29. These data on the effects of practise can be obtained from Dr. F. L. Wells.

(28) Page 34. See Shen '25 and Freeman '28 for criticisms of the Merrimen and Tallman studies respectively. In 1906, Heymans and Wiersma collected data on parent-child resemblances in Holland, which Elderton correlated with results approximating the above ratios. In 1907, Schuster and Elderton measured the resemblance between father and son, and brother and brother in ability to pass the Oxford B.A. examination, and obtained a coefficient of .312 for the former and .405 for the latter. In 1928 Banker used scholastic records extending over a period of more than 65 years, and with the SD's obtained from these, computed the following correlations: Mothers and Daughters .361 $\pm$ .071; Fathers and Sons .517 $\pm$ .066; Sisters and Sisters .138 $\pm$ .072; Brothers and Brothers .660 $\pm$ .054. (See pp. 62–63 of the 28th Yrbk. of the Carnegie Institution of Washington, 1929.) See also Burt '12–'13; McCommas '14; Spearman '14–'15, Downey '18; Holmes '21 (Chap. V); Gates '21; Poyer '21; Elderton '23; Peters '25; Gun '28; and Popenoe '29. For annotated references on sibling and twin resemblance studies seek Burks '28 (b), pp. 252–264. For other and recent studies on twins, see Lauterbach '24; Hildreth '25; Averill and Mueller '25; Wingfield '28; Holzinger '29; Freeman '29; Waardenburg '29 and Wilson and Wolfsohn '29.

(29) Page 35. Dr. C. B. Davenport, in a personal interview said: "By crossing two individuals from two races, each with clearly defined and different features, much can be learned about the heritability of traits: the offspring can be compared first with one parent, then with another, and the distribution of traits in the offspring can be analyzed." In this connection, see studies on the equating of test scores and amount of racial admixture: Ferguson '16; Reuter '17; Holmes '21, (Chap. XI); Garth '21 (b), '22, '23 and '27; Peterson '23; Gregg '25; Paschal and Sullivan '25; Herskovits '26 and '27; Hankins '26; Hunter and Sommermier '22; Garth and Garrett '28 and Terry '29.

(30) Page 36. See early works on race, such as Tylor's "Anthropology", Le Bon's "Psychology of Peoples" and Deniker's "The Races of Man". For later works, see Thomas '12; Kirkpatrick '26; Willey and Herskovits '27; Boas '27 and '28; Hsaio '29; Brigham '23 and Hankins '26. Estabrooks '28 (a) denies that race depends on mental as apart from physical traits.

(31) Page 37. See Burks '29, where the necessity for working with dispersion in human ability, rather than with mean levels, is pointed out. This was also stressed by Dr. David Wechsler during a personal interview with the writer.

Pintner '23 in comparing American and foreign children on language and non-language tests, and Pyle '18, testing white, Indian, Negro and Chinese subjects, both estimate dissimilarity thus, although the latter found that differences between American rural and city children were greater than were Chinese-American differences.

Arlitt '21, although she found that children of the same social status but of different race were more alike than those from the extremes of social status in the same race, nevertheless found a real difference between the proportion of dull to bright children in the white group, and in the Negro and Italian groups; she also found that the very inferior white groups were slightly above the inferior Negroes and Italians. These differences she attributes to race.

Boas '28 says: "If there are 50 per cent of an European population who have a brain weight of more than, let us say 1500 grams, there may be only 20 per cent of Negroes of the same class. There are 30 per cent of the large brained Europeans cannot be matched by any corresponding group of Negroes. It is justifiable to compare races from this point of view." (p. 36) This general procedure is regarded unfavorably, however, by Peterson '23b.

(32) Page 37. See Bere '24, p. 41 and Blackwood '27, p. 93. See also Kellor's study on the emotions of Negresses, referred to by Marett, p. 91 and A. L. Crane's study

on Negro fear reactions. (Arch. Psychol. 63) Studies of temperamental differences between races have been made by Odum '10; Davenport and Craytor '23; McFadden and Dashiell '23; Bond '26; Hurlock '26; Garth and Barnard '27; and P. C. Young '29. The need for tests of temperament for inter-racial use has been stressed by Drs. T. R. Garth (personal interview) and O. Klineberg (9th International Congress of Psychol.). W. MacDougall and A. M. Carr-Saunders think that the clearest racial differences are temperamental. (Ref. Garth '25)

(33) Page 38. In order to prove whether the speed element makes a test discriminate unfairly, subjects should be ranked in the order of their scores, and then retested on different forms of the same test, with the original time either halved or doubled. This was done on the Dearborn Group Intelligence test, and it was found that the relative speeds held the same position each time the test was given. (Personal interview with Dr. D. H. Latshaw).

Freeman '26, (p. 252) by an ingenious experiment, found that the Army Alpha was mostly a speed test. Freeman also found a correlation of .63 between speed and power. Garrison '29 obtained correlations between speed tests (such as card sorting, cancellation) and higher mental ability scores (Otis Self Administering) varying from .22 to .34. Others who have worked on speed are: Farnsworth, Seashore, Clark, Gilbert, Tinker, Hunsicker, Longstaff, Porter, Sisk, F. C. Walters and F. S. Freeman. McFarland has ably summarized studies on speed and mental ability. (Psychol. Bull. 1928). Of great interest in this connection are the studies of Travis and Rounds correlating speed of reflex action with intelligence. (See Appendix II, Page 277.)

(34) Page 38. Personal interviews. For discussions of the meaning of perspective in pictorial tests for primitive peoples, see Blackwood '27, pp. 29 and 68. See also the list of universal test elements — page 44 of this volume.

(35) Page 39. Garth has tested Indians with mental fatigue tests, association and color preference tests and the Seashore test, as well as with regular tests. For a summary of work on Indian testing up to 1927, see Blackwood '27, pp. 108–110. For further examples of the way culture differences vitiate test items, see Pintner and Keller '22 and Hunter and Sommermeir '22. Jamieson and Sandiford have studied Southern Ontario Indians, and conclude that their subjects are handicapped by language and "socio-economic status".

(36) Page 39. For summaries of attempts to evaluate the effects of language and social status in racial comparisons, see Klineberg '28, pp. 6–10 and Burks '28, pp. 291–296. See also Kirkpatrick '26 for a survey of studies dealing with the intelligence of immigrants from overseas, including an examination of the influence of environment on test scores.

(37) Page 39. See Appendix II, 5h, page 277. Dr. M. Mead suggests in this connection that some tribes (e. g. the Samoans) are more interested in words, are more "word conscious", than others. See also E. Sapir's work on primitive languages. (N. Y.: Harcourt Brace & Co.)

(38) Page 40. See for instance Walcott '20; Colvin and Allen '23; Wang '26 and Mead '27. Of great value in this connection is D. W. Seago's analysis of language factors in intelligence tests in "Mental Measurement Monographs" 1926.

(39) Page 41. Ninth International Congress of Psychology, Conference on Race Differences. Dr. Kimball Young, commenting on Mr. Long's paper, stressed the need for an experimental set-up in which Negro and white children should be put in the same situation. See also Pressey and Shively '19, who show up the probable unfairness to Negroes of the test material now being used. See also A. H. Arlitt's communication on the influence of race and social status on the I.Q. (Proc. Am. Psychol. Assoc. 1920).

(40) Page 42. "Nothing in the technique of intelligence tests .... warrants any comparison between the intelligence of various races." (H. C. Link, Board of Educ. Pamphlet No. 44.) See also Oldham: "In the present state of our knowledge we have no means of determining how far observable differences may be innate or how far they are due to native capacity being stimulated or hindered by circumstances." (p. 75) Herskovits '26 criticizes Ferguson's "The Psychology or the Negro", (which purports to prove white superiority) by pointing out, for one thing, that northern Negroes were equal if not superior to southern poor whites, and far above southern Negroes. The same writer

in "The American Mercury" 1925 concludes: "The basic hypothesis of white superiority is to be gravely doubted." See also Symonds '24, Estabrooks '28 and Boas '21. For references on Negro-white comparisons, not already mentioned, see footnote 66.

(41) Page 42. Visual acuity was tested by Snellen's Haken, and Guillery's method. Sensibility to differences of brightness were tested by Masson's disc rotated on a color wheel. Holmgren's wools and the Meyer experiment were used to test color vision; visual space perception was tested by various methods, such as Herring's fall experiment, double images, the Muller-Lyer Illusion, the Bradley-Martin "Pseudoptics", etc. Auditory acuity was tested by Runne's clock and Politzer's Hörmesser. Rivers says that Cohn devised some modifications of Snellen's Haken, making the figures more universal. Rivers also suggests that Guillery's method be modified for ethnological testing.

(42) Page 43. Taken from a book review by Oliver La Farge, N. Y. Herald Tribune, March 30, 1930. That achievement is not an accurate gauge of ability is now generally acknowledged. Woodworth '10 says: "The civilization possessed by a generation cannot be used as a measure of that generation. A comparison of the rate of progress of different groups might serve as a measure of intelligence." Boas '11 protests against the assumption that the mental status of a people is measured by the difference between its social status and our own, and cites different conditions which may account for the slow progress of some races. Nutting '26 demonstrates that no conclusions can yet be drawn with regard to racial inferiority on these grounds, because achievement is too influenced by environment to "furnish a good criterion for fundamental intellectual ability". See also Wallis '26, whose theme is that "culture is casually, not causally associated with race" because "the specific biological type appears to be only a chance carrier of culture" and a "homogeneous culture is carried by a heterogeneous type."

The other side of the argument is presented by Thorndike: "Two races need not be equally gifted because each is equally well adapted to its environment, if the second race has by superior enterprise sought out or created a more exacting but also more remunerative environment. The Bushman may count all that he needs to count, but to put oneself in a position that needs algebra and the calculus may itself be a symptom of superiority. So the complex of qualities which is called enterprise remains largely untouched by the psychologist's tests. The very fact that a certain test seems to be unfair to the Bushman may be evidence of his inferiority." (Educational Psychology. III. p. 221.)

(43) Page 43. This list was obtained from Dodd's monograph and from a personal interview with Prof. F. Maxfield. It has been pointed out by Dr. Margaret Mead (personal interview) that even sense discriminations may not be testable as innate qualities. She cites the case of Manus children in the Admiralty Islands, who can keep a tune before they are six, but then begin to sing in the monotone used by their elders. In the same way, certain primitive people classify colors together such as blue and green, or yellow and green. They have been trained to consider what we call separate colors as merely different shades of the same color. That geometric symbols are not necessarily universal was also pointed out by Dr. Sapir at the joint meeting of the Psychological and Anthropological Sections of the British Association for the Advancement of Science, Aug. 11, 1924.

(44) Page 43. Personal interviews with Drs. F. N. Maxfield, H. A. Ruger, C. Brigham, R. S. Woodworth, R. Pintner, T. R. Garth, G. O. Ferguson, J. L. Graham and B. M. Minogue. Personal letter, Dr. L. D. Anderson.

(45) Page 43. Dr. M. Viteles, although theoretically favoring the development of a "universal test", believes that it would be extremely difficult, if at all possible, to devise one for this very reason. (Letter to writers.) Likewise Dr. D. Wechsler thinks that while it might be possible to test the simpler processes (sensory and psychomotor) universally, it would be practically impossible to test the higher processes since they must be measured by complicated symbols which differ with different social and cultural groups. Thus any test of the higher processes probably tests both native capacity to use the symbol (say language or mathematical signs) and also the degree of familiarity with it. (Letter to writers) Dr. H. E. Garrett thinks that a truly "universal" test would have to be too simplified to be significant. Comparisons between civilized nations with different languages and cultures might be useful however. Dr. Mead was also pessimistic, and thought that nothing interesting would be left to measure, once everything with a taint

of localization was removed from test content. (Personal interviews. See also footnotes 52 and 54.)

(46) Page 43. Personal interviews with Drs. E. L. Thorndike, J. L. Stenquist, W. Healy, A. Bronner, G. H. Kent, D. Shakow, F. Boas, H. T. Potter, S. J. Tulchin and M. Shimberg. Dr. V. A. Jones believes that "until great improvement has been made in eliminating the effect of the social inheritance upon the medium used in measuring native abilities" a universal standard cannot be established. (Personal letter) See also Dunlap '27: "The more generally applicable the tests, the less useful they are."

(47) Page 44. For instance, some psychologists believe analogies to be basic: others, completion of complex forms; and still others consider vocabulary or mathematical problems fundamental for testing intelligence.

(48) Page 44. For instance, the test whose answer depends on the bare fact of having lived in the word is a test which favors the dull child. (Ref. Miss Minogue, personal interview, and Jones '26) Another illustration of the use of item analysis is the weight discrimination test in the Stanford Binet. If carefully analyzed, this is found to test merely the ability to grasp the situation (in low levels of intelligence), because the weight differences are so slight that even some very intelligent people cannot pass the test. To make the test actually one of weight discrimination, the range of weight should be increased. (Ref. Dr. Brigham, personal interview) In many laboratories, the errors made on tests are being studied in order to distinguish ambiguous or worthless situations from those which discriminate between subjects good and poor in the criteria. (Ref. B. Johnson on the Omitted Letter Test, at the Johns Hopkins Psychological Laboratory) Item analysis can be made to serve many purposes, such as to discover the difficulty of items in order to score fairly, or to find which items are most "coachable", etc. References: V. A. Jones' 26; Blackwood '27 (for non-statistical analysis of items in the Dodd and Squires Scales) Dodd '26, p. 17 and Kirkpatrick '26 (for analysis of Beta and Illinois Examination sub-tests). See also studies of S. A. Hamid and L. Vincent on qualitative analysis of test items, the recent work of Orleans, Cleeton and Symonds on Difficulty, and that of Brigham, Stalnaker and Abelson on college aptitude sub-tests.

(49) Page 44. Prof. E. L. Thorndike told this to one of the writers in a personal interview. He added that presumably checks on the validity of such translation of the Binet would be made by using individuals who were bilingual and familiar with two "cultures" or fields of experience. See also the project planned for development in the Psychological Laboratory of Princeton University, (under the auspices of the National Research Council) whereby a verbal scale of the Binet type would be developed, suitable for ready translation into different languages, for universal use.

(50) Page 44. Personal interviews with Drs. F. Boas, F. N. Maxfield, and L. S. Hollingworth. The last qualifies this statement by saying that "nothing would perhaps be ideal save to transplant a whole generation of each people from its own to the other's culture, and note results." (Letter to writers) Dr. A. L. Rogers believed concrete intelligence might conceivably be measured universally provided there were concrete expressions which were universal; for example, cats-cradles might offer an interesting source of material since they were a form of play so widely distributed in the world. (See Haddon '11). In this connection, Dr. H. Hart wrote that tests based on native cultures would be "more apt to reflect degree of cultural development . . . than degree of innate mental capacity." He proposes two approaches to this problem, which might furnish comparable results: 1. Testing infants of primitive tribes with the psychometric tests used on American babies. 2. Comparative study of brain anatomy. (See paper by M. McGraw on testing Negro and white babies under a year old. Spring meeting: New York, Chapter, A.P.A., 1930)

Dr. T. R. Garth told the writer he was studying the possibility of making tests for Indians out of the Indian sign language. Dr. C. Wissler commented that this was, after all, an arbitrary and highly specialized language, just as peculiar to its own group as was any other. General Hugh Scott, in a letter to one of the writers, comments on the great difficulty involved in using the Indian sign language. He does not believe any Indians, except perhaps a few "men of olden times who are still alive", know the sign language, and doubts if any examiner could be found with sufficient skill to give and receive answers. (See also the communication by D. Katz, 9th Internat'l. Congress of Psychol.)

(51) Page 44. Personal interviews with Drs. C. B. Davenport and C. Brigham. The latter stresses the need for investigating different forms of pictorial material; for instance does a pictorial analogy test the same thing as a verbal analogy? In other words, can the relations between the two dimensional representation of objects be as readily deduced as the relations between their verbal representation? Dr. W. Healy thinks that tests for concrete situations might be most worthwhile for inter-racial measurement. In a letter to the writers he said that the shadow item in his Pictorial Completion test was good, but again emphasized the dangers embodied in trying to make pictorial representation universal. (See page 38.) He concluded: "I wonder if someone could not get up some actual miniature representation material that would be far more easily recognized . . than pictorial material."

As regards tests which involve drawing with a pencil, Mrs. R. W. Earle (personal interview) said that, from her experience in testing low levels of intelligence with the Stanford Binet, she had found that failure on the drawings of the square and the diamond (ages four and seven respectively) were among the most diagnostic performances. Dr. Goodenough has written (personal letter) that a combination of pictorial and drawing tests, together with the mazes and formboards, would probably be better than any one of these taken singly, but that actual comparative investigation must be made before adequate weighting systems can be worked out. She mentioned an extremely interesting collection of children's drawings at Leipzig, which was made by Lamprecht at the close of the nineteenth century. This collection, which is probably the most extensive that has been made, includes drawings from children of primitive races. As Dr. Goodenough points out, many of these racial groups have undoubtedly become more civilized since the time the drawings were made, and therefore a comparison of present-day drawings with these earlier ones "would go far to ascertain how much and in what ways the characteristic features of the drawings have been altered by the changing cultural standards of the groups." Ivanoff has examined a small proportion of the drawings, but to Dr. Goodenough's knowledge no other work has been attempted on them. (See also the report of Decroly's communication in the Proc. of the Ninth Internat'l. Congress of Psychol., 1929.)

(52) Page 44. Dr. H. E. Garrett believes that formboards would be novel to differing groups of people to practically the same extent. Dr. G. O. Ferguson, Jr. writes that he thought the Ferguson Series of Formboards would be useful in examining Indians and South Sea Islanders. The greatest need is for a satisfactory method of scoring. (See Herrick '21 for an account of using the Goddard formboard on foreign children.) Dr. G. H. Kent, although she does not believe a universally applicable test is possible, considers that her graded series of formboards comes nearer to universality than any other test. She emphasizes, however, that this is only "comparative" universality — a concept which must be used loosely because actually there is no such thing as "universality".

(53) Page 45. Personal interviews with Drs. D. H. F. Latshaw and A. Otis.

(54) Page 45. Personal interviews with Drs. A. Bronner, J. L. Graham, L. S. Hollingworth, H. E. Garrett, J. S. Orleans, H. O. Ruger, J. Peterson. Prof. Terman, in a letter to one of the writers, says that he is unable to tell how valid our best performance tests are for primitive subjects. He continues: "The anthropologists are apparently nearly all convinced that any kind of tests we might work out in a civilized country would be utterly unfair to apply with primitive subjects. My guess is they are mistaken, but I haven't the proof. I am inclined to think that tests of the effects of practice could be worked out which would be decidedly valuable for inter-racial comparisons."

(55) Page 45. Prof. Woodworth ('29 and personal interview) thinks that in the case of sensori-motor tests, the only way to get a controlled group is to practise every individual up to the limit of his capacity, and then test. Dr. Mead (in a personal interview) said that this could not be done in certain cases, viz: Even if an American adult were practised to the limit of his tapping ability, he could not be fairly compared with a Manus native who had become an expert drummer. The wrist muscles of the latter would have been developed since childhood in a particular tapping motion, and the former could never hope to equal him in speed. Again, in a "recognition of objects" test, we might practise ourselves to the limit of our capacities, and yet a people with a non-verbal language would be immeasurably superior, having been trained to watch for objects as we watch for words. The same holds true of string puzzles and bead

stringing tests. In certain tribes, the smallest children become experts at these occupations. These observations raise difficulties for Dodd's definition of intelligence as "ability to achieve after comparable training." (See also Need 4h. page 276 of this volume.)

Klineberg '28, p. 90, suggests the possibility of a qualitative difference between the intelligence of various groups, such that one group may excel in ability to deal successfully with new situations, the other in ability to profit from experience. He gave the Healy Puzzle A twice in rapid succession to whites and Negroes, and found a distinct superiority in the latter's actual gain, but very slight relative improvement. He therefore thinks that there is little difference between the two groups in ability to profit from experience. Pyle '16 on the other hand, after testing the building up of motor coordinations by means of a "learning apparatus", concluded that Negro children have only three fourths to four fifths the learning capacity of white children. Peterson and Barlow '28 studied the effects of practice upon individual differences in initial learning scores, and found that in digit-symbol substitution, card sorting, mental multiplication and achievement test performance, the subjects tended to keep the same relative position at different stages of practice. The authors consider the probability that practice decreases individual differences in simple psychomotor operations, but increases them in performances which tax the higher mental processes. Ruch '25 found that the correlation between initial and final status in learning is greatest for tasks which appear to involve the "higher" (abstract) mental processes, and lowest for those depending on simple perceptive processes and manual dexterity.

See also the work of B. T. Baldwin on the learning of delinquent adolescent girls; O. J. Johnson on test scores and learning; F. Goodenough on learning and accomplishment; B. Johnson, A. Gesell, L. H. Meek, K. Shaw, E. McGinnis, J. A. Kirkwood, etc. on the learning of young children; J. Peterson, Lanier and Walker on white vs. Negro learning; E. L. Thorndike and H. E. Garrett on adult learning; F. T. Wilson on the learning of bright and dull children; W. R. Atkinson on mechanical speeds and learning, and W. H. Pyle on the general nature of the learning capacity.

(56) Page 46. This information was obtained through correspondence. Dr. Graham emphasized that his learning test (see page 84) was adapted to high levels of intelligence (a level requiring "abstract, symbolic and relational reactions") but could also be adapted to lower levels. He showed further how difficult it was to score a learning test, because learning is largely of the unit and not of the parts, and hence the separate tests could only be scored arbitrarily.

(57) Page 46. Dr. Maxfield, after working with the Squires test, found that the piled cubes were "cumbersome as test material, and frequently came to pieces and required repair." One of his graduate students, Mr. Hoffman, made drawings of the piled cubes on square cards, and adapted Squires' method to this material, though no attempt was made to establish equivalence with Squires' material. Maxfield thinks that Hoffman's preliminary work has shown that the drawings can be used instead of the cubes in a non-verbal test situation. Maxfield found the tachistoscope in the dot-pattern test (Squires) very unsatisfactory, and suggested the use of a standard tachistoscope in its stead. He also thought it unfortunate that such common types of non-verbal tests as formboards and mazes were not included in Squires' test since some of them (notably the Seguin and Ferguson formboards and the Arma Beta Maze) could be given in pantomime. (Reference: letter to the writers)

(58) Page 46. Peterson obtained high correlations between time and errors and therefore disagrees with Klineberg '28 who stresses speed as an independent factor. Dodd '26 makes a point of urging that time limits be dispensed with, thus testing nothing but "power", and "avoiding the speed factor, except where it proves by correlation to have predictive value."

(59) Page 47. Dr. Mead told one of the writers that she would not attempt to construct a test for primitive tribes before she had lived among them for at least five years.

(60) Page 48. "It should be understood (however) that the degree of multiplicity required varies considerably for different tests. Tests based primarily upon school knowledge do not have to be standardized separately for the two sexes, but many of our performance tests should have separate norms for boys and girls. The maze test, for example, has been shown by Young to give appreciably different results for male and female subects, at any age level."— (Personal letter to the writers.) See also the study

of Kemal '28 on sex differences, which proposes that when boys and girls differ too much they should have separate scales.

(61) Page 49. In this connection it is pertinent to mention the suggestion of Porteus and Babcock '26 that if mechanics, electricians or farmers had constructed a test and scored psychologists on it after standardizing it on themselves, the tables would be reversed as far as the I.Q's. go. "By our present general ability tests, the only superior adults are school masters and psychologists."— p. 190. Kirkpatrick '26 says: "It may be that ditch diggers would approach the achievement of college professors in a test of mechanical ability."— p. 33.

(62) Page 49. See Thorndike, Chap. 16, '21 and et al '26. "In measuring a person's general status in intelligence, and in inferring therefrom what his rank in native intellectual capacity in general is, what we do is to test him with a fair sampling of data and operations. If his opportunities of training in respect to these have been inferior or superior to the group with whom he is to be compared, we make the necessary allowance. This sampling should be wide enough, and its various components should be easily enough weighted, so that the resulting judgment should be about his general status and general capacity— if we are to claim that it is."—'21, p. 126.

Drs. R. S. Woodworth and G. H. Kent also support this proposal. The latter says: "It is important to have on hand a great variety of tests, in order to have something that is fairly applicable to any given subject of whatever age, race or training. Some tests are much more widely applicable than others, but it should be recognized that all of them are of limited applicability."— (Letter to the writers.) See also Vabalas Gudaitis '29.

(63) Page 50. Dodd '26, Blackwood '27, Porteus '17, Macrone '28, Stoneman '29 and Peterson and Lanier '29.

(64) Page 50. In statistical terminology, a group is a "sampling" taken from a "population" which includes all the individuals who fulfill the terms of the category which has been selected, say, race, schooling, or social class. Every group comparison involves different homogeneities and heterogeneities. For instance, G. H. Estabrooks '28 (a) writes: "If we wish to compare gross scores (of two races), we must confine ourselves to the same cultural groups and probably to the same social strata within those groups."

Dr. V. A. Jones says in a letter to the writers: "For theoretical purposes inter-racial comparisons may be important when satisfactory measuring instruments are devised; however, even then a sharp distinction must be made between comparisons made among racial groups in a heterogeneous population such as in the United States, and comparisons based on more pure and representative samples of racial stocks in relatively homogeneous homelands."

See also Antipoff '28, who emphasizes the necessity of carefully accounting for heterogeneity whenever it creeps in. The criteria she suggests for dividing up homogeneous groups are: economic, ethnographic, sex, pedagogic, professional and "caracterologique". She urges psychological research to create precise "social indices", and cites Terman and Descoeudres as giving results according to different social milieux.

(65) Page 52. From the introduction to Part I of the 27th Yrb. Nat'l. Soc. Stud. Educ. by L. M. Terman.— p. 3.

PART III.

Introduction

Bibliography to be found on Page 63.

(66) Page 61. Prof. Maxfield thinks it is highly desirable to establish procedures and secure norms in a strictly non-verbal situation. (letter to writers.) Boody '26 was forced to use pantomime in her work with immigrant children, because, as she says: "Non-language tests so often mean non-language on the part of the student, but much language on the part of the experimenter". Squires '26 points out that absolute standardization of pantomime instructions is far more difficult to attain than is verbal instruction. He suggests moving pictures as a possible solution, but admits the almost insuperable difficulty of exact reproduction. Pantomimic directions require more time, and are less reliable than verbal ones because they have fewer elements.

The writers are aware that to introduce a new variation into the administration of any test (namely, pantomime instructions) necessitates re-standardizing the test under these new conditions.

(67) Page 61. Modern psychologists are realizing that mental functions are not single activities, since there is a wide variety of reaction in any situation. Garrett and Lemmon ('24) say: "Indications that even if we do not know exactly what a test measures, at least we realise it, are seen in the tendency of present-day workers to call a test by some descriptive title rather than in terms of some more or less vaguely defined mental function which it may be thought of as "measuring," and so we have cancellation tests, opposites tests, completion tests, etc., instead of tests of perception, association or reasoning." (See also page 10). The partial correlation technique has thrown light on certain relatively independent factors involved in a test, however.

(68) Page 61. This study might have been extended to include a historical sketch of each test, its original construction, subsequent modifications, various standardizations, its employment in various researches, the merit it seems to possess most consistently, its appearance in certain batteries or scales and its present status.

(69) Page 62. Those responses which involve drawing lines, symbols or pictures on paper are here designated as "non-verbal;" "manipulative" means handling three dimensional material, such as blocks or beads.

(70) Page 62. Dr. David Wechsler has pointed out (personal interview) the many unmeasurable factors that enter into studies of emotion as seen in respiration, blood pressure, capillary pulse, reaction time, etc. He thinks, however, that the psychogalvanic reflex (as an index of emotional intensity) does seem to lend itself to quantitative analysis, and that a point has been reached where a number of sensory stimuli can be standardized. He stresses two difficulties, however: first, the provoking of "real" emotions in the laboratory; second, the equating of stimulus words in testing other nationalities. See also the work of W. Whately Smith, who stresses the difficulty of comparing the reactions of two people on the galvanometer, and the account of Dr. Carney Landis' work, (page 275).

(71) Page 62. Some outstanding group non-verbal tests for pre-school, kindergarten and primary grades are: Cole Vincent, Pintner Cunningham, Rhode Island Intelligence, Dearborn Group (Series I), Haggerty Delta, Detroit First Grade, Kingsbury Primary Group, Otis Group, Pressey Primer, Trabue and Stockbridge's Mentimeter and the Town Picture Game.

(72) Page 62. It is certain that infant and pre-school child psychology offers a vast and fruitful field for research in tests free from formally learned elements. Sensori-motor co-ordination as signs of "intelligence," and the development of social behaviour patterns are phenomena demanding (and receiving) investigation. The whole topic of growth and learning will probably yield more facts about heredity than any other non-biological science.

(73) Page 62. The line between a laboratory experimental device and a test is drawn thus by Whipple in his Manual:

"The primary difference between the research experiment and the test experiment is really one of aim. The test has a diagnostic rather than a theoretical aim; its purpose is not to discover new facts,—or laws for the science of psychology—but to analyze, measure and rank the status of the efficiency of traits and capacities in the individual under examination. It studies mental performance rather than mental content."

(74) Page 62. Puzzles are a kind of test material which is too much dependent upon chance for successful solution. Puzzles calling for patient trial and error yield the most uniform results, but even so, their true significance in the development of tests is uncertain. Most workers have discarded them in favor of more stable materials (See Dearborn's "The element of surprise is not desirable in a test."); but Ruger is still interested in them, believing that they have possibilities of future development. It is for this reason, rather than for past contributions, that certain puzzle tests are included in this compilation.

(75) Page 63. We wish that test constructors would recognize their earlier important but now out-of-date attempts, and repudiate what no longer holds good with as much readiness as does one test author who says:

"Every time I discarded a model, I did it for a definite reason. Each time I was thus able to eliminate at least one defect. It is because of these unsuccessful attempts that I was able to make my final model as nearly perfect as it is, so I can claim that this model represents all the strongest features of the various discarded models."

(76) Page 63. Advice on reporting test data has been given in many excellent articles, notably by Otis, Ruch '25, and Smith and Wright '28. If some standard form were adopted, the confusing variability of titles, research techniques and methods of presenting results would cease. The problem of nomenclature may be illustrated by citing the case of a test which, according to its author, is called the "Healy Construction Test A". This has been named by five different and authoritative sources: "Healy Frame Test," "Healy Puzzle A," "Healy A," "Construction Puzzle A," "and "Healy-Fernald Formboard No. 1." Statistical data offer different norms, depending on which revision of a test is used, and this is also left to the guesswork of the reader. If, as Pintner and Paterson '17 suggest, workers would publish their results in such form that the new might be added to the earlier data of previous workers, much labor would be saved. We may be pardoned the suggestion that a committee be appointed similar to the one which drew up the article "Instructions in regard to Preparation of Manuscript (Psychol. Bull. 26, 1929, 57-63) urging a standard form of test reporting .

Chapter II. Statistical Bases of Tests Interpretation.

Bibliography to be found on page 75.

(77) Page 72. "S. D." means "standard deviation," which is a method of indicating how many points above and below the average, 68.26% of the cases will fall. See Garrett '26. p. 26ff.

The importance of knowing about the age range to which a test is applicable is realized when, as Squires '26 points out, the same test is seen to measure different things at different ages. Terman '21 says: "No one type of test will measure intelligence equally well at all levels. What it tests in lower levels may be one thing, e. g. eye-hand coordination; what it tests at another level may be a subtle reaction to relationships of forms, involving abstract ideas." Burks and Kelley '28 point out that "data collected by different experimenters may fail to agree simply because the range of the subjects is different in every case."

(78) Page 74. Another measure of rating in terms of the variability of a group is the "T-score" described by McCall ('22, p. 195). The T-score is based on the mean or average of the scores made by a group of unselected twelve-year-olds and also the standard deviation from this mean. By application of the statistical formula to these two measures, the T-score can be derived, and thus a finer gradation can be achieved than is possible by the use of the S. D. alone.

(79) Page 74. "Two plus three has so often totaled five, and two times three so commonly yielded six, that we have assumed test scores may with entire propriety be added, subtracted, multiplied and divided. They seldom can. Test devisers have apparently been quite successful in obtaining test score units, which are substantially equal and can be added and subtracted, but they have failed quite signally in determining reasonable zero-points so that the product or quotient technique rests upon shifting ground."

Toops ('26, p. 137) offers positive statistical guidance: "Test scores may be weighted before combining them into a composite scale score, which will predict the criterion better than if weighting were not resorted to. The maximum correlation of the weighted composite with the criterion is secured by weighing each test, proportional to its partial regression coefficient. The partial regression technique is so laborious that it is known to but a few."

PART IV

Summary, Recommendations and Conclusions.

Bibliography to be found on pages 255–256.

(80) Page 246. It seems that, despite language and culture difficulties, inter-racial comparisons by means of tests do yield significant results. The mere fact that Yeung '21, Darsie '25 and Sandiford and Kerr '26 found that Chinese and Japanese children ranked with Americans (chiefly on non-verbal tests, it is true), would go to prove that these

difficulties may be overestimated. Indeed, it seems as though many investigations had actually revealed race differences in intelligence and sensori-motor functions. Among these, the following are cited: Rigg '28 finds low Italian scores as compared with Jewish ones, and concludes that the slight language handicap is insufficient to explain them. Likewise Graham '26 finds that a group of Jews is superior to a group of Italians even allowing for language handicap. Klineberg '28, after giving a resumé of the literature on environmental differences in relation to race differences, proceeds to find "real" differences in speed between Indians and whites. Seago and Koldin '25 conclude that Jews are superior to Italians, and "the difference is not essentially one of language handicap." Peterson and Lanier '29 found that white adults were reliably superior to Negro adults in group intelligence tests, and Peterson '28 obtained differences between whites and Negroes on his Rational Learning Test, although it depends very little upon schooling or social milieu. Again, Murdoch '20 reports test results for Hebrews, Italians and Negroes (all considered as having no language handicap) as follows: The Hebrew median was surpassed by 57.7%of the Americans, 30.3% of the Negroes, and 15.5% of the Italians. Mayo '13 found that only 29% of Negroes exceeded the white median of high school marks. For other Negro comparisons not yet referred to, see M. N. Work's "Bibliography of the Negro in Africa and America" (N.Y.: H. W. Wilson Co. 1928. Pp. 698), pages 570–576. Besides this see Schwegler and Winn '20, the Negro Yearbook for 1925-26, Strachan '26, Viteles '28, Price '29, L. Graham '30 and C. S. Johnson '30. For summaries of racial studies in general, see Bruner '14, Woodworth '16, K. Young '22, Garth '25, Goodenough '26, Kirkpatrick '26, Klineberg '28, Yoder '28 and L. H. Lanier's communication to the 9th International Congress of Psychology.

In the realm of performance, sensori-motor and special abilities, Davenport '29 reports differences between Negroes and whites in Jamaica. The former were superior in discriminating minute time differences and in capacity for rhythm, while the whites were superior in visualizing forms (copying geometric figures), seeing absurdities in test sentences, fitting blocks into a formboard, and drawing a man—"whatever innate ability that presupposes." They were also superior in "closeness of observation, ability to visualize, and to trace with a pencil." See also Lenoire '25 on the superiority of the Negro to the white in rhythm and tonal memory.

(81) Page 246. See Porteus and Babcock '26. "The particular task of racial psychology is to . . . attempt to separate essential differences, or those which depend on innate capacities, from the indirect or accidental, which are due to environment." See also Garth '25: "Race psychology, as a problem, is an effort to . . . . measure racial mental characters and behavior, which are determined through racially inherited nervous mechanisms, involving cautiousness in taking into account social status and all facts of nurture and education, until we are able to perfect tests that will measure abilities in spite of the presence of nurture."

(82) Page 246. Ref. Dr. Clark Wissler (Personal interview). The fact that two cultures may appear alike on the surface, yet differ radically underneath, is pointed out by Wallis '26, who goes on to say:

"The mind does not develop in a vacuum. It develops only in a world in which organism acts on environment, and environment on organism. In developing, it takes on something of the complexion of the environment to mental life. Given a common environment, we have a common basis for comparison."— p. 316

See also Mr. Post's report on international anthropological measurements. (Carnegie Institution of Washington)

(83) Page 249. This was urged by Dr. Dunlap (personal interview), who said that a central body which would serve to coordinate research workers is needed in every branch of experimental science. See the Engineers Yearbook.

(84) Page 250. Practically the same words are used by Goodenough, in a paper read at the Toronto Conference on Child Welfare, N. R. C. 1929.

(85) Page 250. Written to the authors.

(86) Page 251. From Dearborn et al '23, p. 38, who were quoting from Pintner-Paterson's "Scale of Performance Tests."

(87) Page 251. The collection referred to is in the Graduate School of Education, Harvard University, Cambridge, Mass. Many other individuals have made collections of tests for their own use, but needless to say, these are by no means complete. e. g. The

Bureau of Educational Tests and Records, New York City; G. Hildreth, Lincoln School, etc.

(88) Page 251. Dr. Myra Shimberg gave valuable suggestions and helped to write this prospectus. It is to be understood that none of the other names quoted as supporting a central agency for tests are in any way connected with this prospectus.

(89) Page 251. Besides earliness of reaction (See Kuo '29), other criteria of "nativeness" have been given by various experimenters—e. g., dependence on maturation rather than on learning, universality, biological utility, and correlation with some structural character.

(90) Page 252. "Intellectual tasks, success in which requires no training, and is influenced by any kind or amount of training, do not exist, at least not in shape to measure appreciable amounts of intellect." Thorndike et al '26, p. 436. See, however, writers such as Stormzand '22, who argue that certain tests are valid measures of general intelligence, and therefore can be used by eugenists for the segregation of morons.

(91) Page 252. A review of studies dealing with group differences is to be found in the chapter entitled "Interpretation of Intelligence Tests" (Freeman '26). Here reasons are given for and against the hypothesis that intelligence test scores are dependent chiefly on native capacity. See also Woolley '26, page 486, for her reasons for thinking that "mental tests constitute a type of measure of the individual in which native endowment is a larger factor than any known standard of measurement." (See also Holmes '29.) Of interest in this connection is Whipple's statement: "We know that the organism arrives at approximate maturity of growth in stature and in many other physical traits in early adolescence; the fact that our test scores indicate the maturing at about the same time of whatever it is we are measuring, like the fact that, regardless of chronological age, the correlation between stature and mental age is high, may very well indicate that our tests are also measuring an intrinsic capacity which matures according to laws of its own and with relatively little influence from the environment." ('22, p. 507)

(92) Page 253. See M. R. Trabue's communication, 9th International Congress of Psychology, 1929.

APPENDIX II

PROSPECTUS FOR A BUREAU OF MENTAL MEASUREMENT

1. A central agency or bureau in the field of mental testing would make a preliminary complete survey of all the institutions and individuals in this and other countries engaged wholly or in part in psychological measurements, including, besides mental tests, measurements of sensori-motor processes, special abilities and emotional, personality and character traits. A special division of the Bureau, co-ordinate with the one here proposed, would have to be established to take care of such work on physical measurements as seems to be related to the work of the Bureau. Material already provided by comprehensive surveys of literature and special investigations such as the Character Education Inquiry (directed by Hartshorne and May) could be used as a starting point. It is to be noted that such researches have always been made sporadically, in only one part of the field; moreover, often no provision is made to keep up the records of such surveys, and consequently they soon become antedated.

II. This agency would establish a relation with these research institutions and individuals such that they would remain in touch with the Bureau, thus keeping its files up to date. The Bureau could, of course, supplement this voluntary information by periodic inquiries, and if it attained sufficient prestige, there would be little difficulty in obtaining answers.

III. It would make a complete compilation of all articles pertaining to mental tests and their administration to various groups, with special references to the needs of those interested in tests from other angles than the psychological or educational.

IV. It would publish a yearly report of its work, references, etc.

V. It would promote high and uniform standards for giving and interpreting tests, so that results from various sources would eventually be fairly comparable. For instance, when tests are not standardized the same way, it is impossible to tell if a mental age of eight means the same thing in all tests giving M. A. scores, and again, it is difficult to liken a group of test results given in percentiles to another using T-scores. If the many hundreds of studies on say, racial and national differences, had been gathered in some standard fashion, these separate pieces of information from many sources could have been made to form a more or less composite picture or at least have been provided with a "common denominator" for all. As it is, with so many different tests under as many sets of directions, given to as many different composite groups, no valid inferences are possible.

VI. The Bureau would act in somewhat the same way as the N. C. M. H. does—disseminating needed material to various institutions, schools and clinics. It ought also to answer all inquiries as to test sources and references. This is necessary so that the most up to date and serviceable data will be collected. For instance, it has been found that certain performance tests are unreliable, and yet they are still being used indiscriminately. Results from their use have little value.

VII. A central agency would make a definite attempt to correlate various pieces of research. In order to do this it would be necessary to have a special staff which would analyze the material and observe it in relation to other studies. This would prevent overlapping, and enable the Bureau to encourage the most needed research. Examples of work now in progress with which a central agency ought to be in touch, have been gathered at random, (chiefly through personal interviews or correspondence) and are given below to illustrate how wide the scope would be:

EXPERIMENTAL MATERIAL

Sensori-motor, structural and performance tests: At Leland Stanford University a raised finger maze is being developed. At the University of Texas, a new pencil maze has been reported. At the University of Oregon, work is afoot on the Stanford motor skills unit, and at the University of Pennsylvania the machine shops are steadily putting out new apparatus for the testing of various motor and other functions. Dr. L. Wagoner is studying motor development with a new apparatus; the Vocational Adjustment Bureau of New York City is working on a team of selective tests for power machine operators, such as paper feeding, tracing, candy packing, pasting, and eye-foot motor coordination. Motor coordination is also being studied at the Johns Hopkins Psychological Laboratory by making subjects trace mazes in time with a metronome. Dr. G. H. Rounds has been correlating the speed of the Achilles tendon reflex with speed of mental operations. (See also Travis '28) Drs. Rand and Ferree are getting a workable battery of functional tests for the human eye, and refining old tests. Dr. Knight Dunlap has developed a new test for color vision (the Nela test).

Tests of emotions: Dr. Carney Landis proposes to put together a test or fundamental procedure which will have "internal consistency", because he finds that psychogalvanic readings have very low correlations with eight or ten other tests of emotion and character. He has succeeded in obtaining items in various tests which seem to hold together as indicators of various personality traits. He hopes eventually to evolve a paper and pencil test which, after having been given to many different groups, (e.g. schizophrenics, morons) will show a personality classification on the basis of a psychograph. At Johns Hopkins, the emotions of the child are being studied by measuring his tensions by means of a rubber tube stylus. His temperamental reactions are gauged by asking him to pick up little metal balls slightly charged with electricity and noting his reactions. His motor control of emotion is diagnosed by means of a specially constructed chair whose back and sides register his every move when emotionally charged objects, such as different kinds of animals, are brought in. (See also Goodenough '29).

Pre-school tests: A new Stanford-Binet is being evolved, consisting of two complete new Scales running at least as low as two years and much higher than the present Stanford-Binet. At the lower end many performance tests are being added, and the entire scale will have more alternate tests than the present Stanford-Binet. Stutsman's dissertation, which is an enlargement of the Merrill Palmer Monograph, is about to appear. At the Universities of Minnesota and Iowa new tests for pre-school children have been reported. Dr. D. Thomas is working out a standard method of recording children's behavior, such as laughter, amount of space covered, etc. (See Child Development Monograph No. 1. Teacher's College, N.Y.) R. E. Atkins reports an object fitting test with pantomime directions, which correlates .70 to .90 with the Binet. (Unpublished Ph.D. Thesis, University of Minnesota)

Pictorial tests: Dr. F. Maxfield is working on a series of simple graded pictorial tests. Dr. L. M. Crabbs has just completed a group test for ages three to seven utilizing the same principle as McCall's Multi-mental Scale but in pictorial form. (This is still in the experimental and unstandardized stage.) At Johns Hopkins, work is afoot on the process of recognizing and selecting geometric figures.

Universal tests: Dr. J. L. Graham is at work on a bead test (see page 84) and Dr. F. Maxfield has two students working with the Squires material. At the Brush Foundation (Cleveland) Dr. L. D. Anderson is developing inter-racial tests for possible use on Negroes and whites. During the summer of 1929 at Cold Spring Harbor, Long Island, Dr. C. B. Davenport and two assistants devised a new bead-stringing test, and a form-color substitution test, based on the Woodworth-Wells model. A new memory span test was also worked out, based on a graded series of nonsense syllables. (See tests on pages 81, 197, and 200.)

It is to be understood that the above list was chosen and classified arbitrarily, and that it represents only a drop in the bucket as compared with the total amount of research in progress. Laboratories of experimental psychology are furnishing valuable information on memory and the learning process; and, besides the invention of new tests, the slow process of standardization is being carried forward by many painstaking workers.

VIII. Finally, the Bureau would attempt to have on record all data already gathered, or in process of being collected in many places, which is not being worked for research purposes. The Bureau itself might eventually work some of these data, or, in any case, be able to point it out to appropriate workers—e. g. candidates for the degree of Ph. D. There are many laboratories and social institutions throughout the country that are well equipped to do the work of human measurement on which eugenic research must be based. By being in touch with all these agencies, the Bureau experts could correlate pieces of research and make needed suggestions. They might also send workers to properly equipped centers to fill in certain gaps, or supply funds to the proper institutions. To establish a special staff of psychologists, together with the laboratories and equipment necessary for eugenic research, would probably lead to unnecessary and expensive duplication of work being done elsewhere. Instead of struggling alone to collect the data it needs, eugenics should utilize the vast amount of material already assembled in many and varied fields of study.

UNWORKED MATERIAL

The present writers know of institutions and individuals who have on hand many tests and records which either are not being used at all, or else can only be applied within narrow limits because no money is available for research. For instance, in the Judge Baker Foundation, thousands of case studies and test results have been collected, but remain in the files unworked. Again, Dr. G. H. Kent has on hand considerable unworked test material, which, if it could be developed, holds valuable possibilities. The same is true of Dr. H. A. Ruger, who has considerable puzzle performance material as yet largely unstandardized. Actual manufacture, or the control of manufacture, is requisite for the permanent production of standard type "fool-proof" test material of this kind. Considerable initial expense would be required to make this material avialable for general testing and research.

Dr. I. S. Wile has a vast amount of pictorial material on file (man and house drawings by children) which could be used in studies of ideational levels at different ages. He has no time to work this himself but would willingly submit his material to the proper research body. Again, Dr. F. L. Wells has worked out an interesting pegboard test, based on a pencil and paper principle—namely, arithmetical rhythmic progression.

Besides these specific examples, there are any number of clinics (e.g. the Witmer Clinic) where practical every day experience leads workers to set up new combinations or adaptations of test material originally worked out in the laboratory. These should not be lost.

A partial list of problems in this field still requiring solution has been collected, again more or less at random, and is here submitted to illustrate the many questions with which a Bureau such as the one described above would occupy itself. (Attention is called to the fact that needs of various kinds have been outlined all through the report, particularly in the section on heredity and environment.)

SUGGESTIONS GATHERED AT RANDOM FOR FUTURE STEPS IN RESEARCH ON THE MEASUREMENT OF INTELLIGENCE.

NEED FOR :

1. **Discussion of mental test findings.**
   a. Derivation of psychological fundamentals from the application of mental tests. Ref. Drs. C. C. Brigham (Personal interview), E. L. Thurstone ('21), H. A. Toops (Communication, 9th Internat'l Congr. Psychol.) and Terman ('24).
   b. "Acute analysis of mental capacities, backed up by statistical examination of the results."— Ref. Freeman ('26).
   c. "Examination of the relationships between many differently labelled, derived, measured, and variously sponsored mental traits . . . . Determination of the independent mental traits, their laws of functioning, and of what adult activities demand of them."— Ref. Kelley ('28).
   d. Study of the learning process.— Ref. Young ('24) and Gates and Taylor ('25).
   e. Study of the difference between inter-test correlations for children and for adults.— Ref. Ruml ('21).
2. **Item analysis.**
   a. Determination of a test's diagnostic value against proper criteria—i. e. the checking of success or failure on the test with actual accomplishment.— Ref. Dodd ('26) and J. O'Connor (Letter).
   b. Study of the inter-relationships between tests in any one field, such as verbal, mathematical or sensory tests.— Ref. Dr. Brigham (Personal interview).
   c. The use of the form of the learning curve as a partial criterion for inclusion or non-inclusion of a given type of performance in a test series. Ref. Goodenough (Conf. Child Dev. Comm. Nat'l Res. Counc. Toronto, 1929).
3. **Standardization and scoring.**
   a. Use of standardized techniques for reporting norms etc. Ref. Burks and Kelley ('28, p. 20.) and Bureau Prospectus, Section V.
   b. Intensive studies on individual tests, to find the best method of administration and scoring.— Ref. Terman ('21) and Colvin ('21).
   c. Measures of group dispersion.— Ref. Burks ('29).
4. **New Tests.**
   a. "Independent measure which does not correlate with the amount of schooling one has had on the one hand, but does give positive evidence of measuring intelligence, as indicated, among other things, by scholastic achievement, on the other hand".— Ref. Freeman ('26, p. 449).
   b. "Tests which call for controlled interplay of disparate mental functions, and yet call for no rich mental content."— Ref. Town ('16) (p. 230. See also inter-racial testing, need 5 e).
   c. "Tests of sustained rational ability where speed is relatively unimportant." Ref. Colvin ('21). Tests without time limits for certain clinical purposes. Ref. Kent and Shakow ('28).
   d. Infant and pre-school tests.
   e. Character tests. Ref. Watson ('26) Colvin ('21), Thurstone ('21) and Haggerty (21). "Thoroughly standardized tests which give objective quantitative measures of special intellectual capacities, or of non-intellectual traits." Ref. Freeman ('26 p. 449). "Tests enabling us to make an all-round examination, besides tests of special ability and temperament." Ref. Porteus and Babcock ('26 p. 189).
   f. Special ability tests. Ref. Terman ('21), Haggerty ('21), Pintner ('21) and Wissler ('29).
   g. Sensori-motor tests.
   h. Learning tests. Development of tests of little used (and therefore little trained) mental functions—e.g. cancellation. Ref. Whitely ('11). (NOTE: This would necessitate studies of the connection between the little and much used mental functions—see Need No. 1 above.)
   i. Non-verbal tests. Tests to measure capacity to learn in other directions than the ability to use words, signs and symbols. Ref. Henmon ('21). Non verbal tests for the higher stages of intelligence. Ref. Squires ('26). "Series of fundamental tests and norms for psychomotor capacity–motor problems the skillful solution of which puts a premium on the individual's keenness of observation, power of intellectual analysis and perceptual discrimination, and his ability to adjust himself to a novel situation and to profit from the memory of former adjustments." Ref. Wallin ('16 p. 354).
   j. Special tests for border line deficients. Ref. Kent and Shakow ('28). (NOTE: The development of such tests is important from the eugenic point of view, because of the danger that morons will marry into normal lines. Ref. Dr. H. Stuart, lecture, Hunter College, New York).
   Tests for psychopathic disturbances.

5. **Inter-group testing.** (Comparisons between groups differing culturally and racially.)
   a. "Clearer definition of the groups studied, and wider knowledge of the background and opportunities of the subjects." Ref. Bere ('24, p. 3). "Analysis of the groups on which tests are standardized, according to such categories as sex, racial and social composition, range, and deviation." Ref. Shimberg ('29, p. 6). "Ultimate determination of norms for national stocks." Ref. Wissler ('29 p. 12).
   b. More outside criteria of intelligence. Ref. Terman ('21), Dodd ('26), Thorndike ('26) and Brigham (Personal interview and '26). "As we broaden our concept of intelligence we must broaden our criteria of intelligence. Teachers' estimates, school marks, etc. are in themselves too narrow." Ref. Pintner ('24).
   c. Research on the rate of growth of different cultural and racial groups.
   d. Research on race classification.
   e. Experimentation with non-verbal tests. Ref. Wissler ('29), Boody ('26) and (4 i) above. A survey of the principles of current intelligence tests to see how many could be switched over into another (non-verbal) medium, e. g. completions, opposites, similarities, analogies, substitutions, etc. Ref. Schwesinger. Investigations of pictorial material. Ref. Brigham—App. I, No. 51.
   f. Studies of fundamental reaction methods for the construction of test devices. Ref. Dodd ('26, p. 14).
   g. Collections of data on peoples of diverse cultures. Dodd ('26) and Blackwood ('27). Research, for instance, to find suitable criteria of social status among foreign peoples for purposes of classification and also to evaluate the effects of race mixture. Ref. Blackwood ('27 p. 34)
   h. Experiments on translating and standardizing verbal tests into other languages and on other peoples. Ref. Drs. Brigham and Thorndike (Personal interviews). Use of individuals who are bilingual and familiar with two "cultures" so as to check the validity of test translations. Ref. Dr. Thorndike, (Personal interview). Experiments in testing different peoples with both performance and language tests, so that intelligence could be estimated on the basis of both language and non-language tests. Thus the factor of language alone might be held constant and the residue be assumed to test something not so immediately dependent on environment. Ref. Schwesinger. Studies on the thought patterns of different types of speech and on the effects of bilingualism. Ref. Blackwood ('27, p. 23 and 26). (NOTE: See for instance the work of Saer, Smith, Murdock and Maddow, and Yoshioka on bilingualism.)
   i. Studies of conditions of immigration of the various racial groups tested in this country. Direct comparison of the immigrant groups with as large as possible a random sampling of the original groups in the original habitat. Ref. Klineberg ('28) and Bere ('24). (NOTE: The fact that Chinese immigrants test higher than Italians in America may be due to rigid selection from China. Ref. Klineberg. It is interesting that in the examination of aliens coming to the United States standards derived from immigrants of the same nationality as the subject are used. Ref. Letter from R. C. Williams., Ass. Surg. Gen. Div. of Sanitary reports and Statistics, Washington, D. C.)
   j. Studies on children in pioneer settlements, who have had few cultural advantages or educational opportunities, but who are the offspring of progressive parents (as vs. canal boat and gypsy stocks in Gordon's study. Ref. Schwesinger).
   k. Intensive study of a single primitive group over a period of time to discover what results such selective factors as the departure of the more enterprising, and the return from civilization of those unfitted to cope with it, might have on the sampling problem. Ref. Blackwood ('27, p. 35.)
6. **Studies on the relation existing between physiological factors and intelligence.**
   a. The relation between speed of neural conduction and I.Q. Studies of individual time responses. Ref. Travis ('28), Rounds (Unpub.) and Dr. Brigham (Personal interviews).
   b. The relation between endocrine development and I.Q.
   c. The relation between basal metabolism and I.Q. Ref. Mrs. R. W. Earle (Personal interview).
   d. Studies to distinguish maturation from environmental influences. "Index of physiological maturity for chronological age in establishing norms and in comparing scores of individuals with norms." Ref. Freeman ('26, p. 293). (NOTE: Freeman goes on to point out that the "rate of intellectual maturing corresponds more closely to physiological maturity than it does to chronological ages." This opens up the whole question of mental development and growth).
   e. Studies on ease of forming conditioned responses.
7. **Studies with tests on mental heredity.**
   a. Further studies of infants, twins and foster children.
   b. Studies of mental traits to discover whether they are independent, and if so, whether they are acquired or hereditary; if the latter, how they behave in racial crossings, whether they are Mendelian factors or "additive complexes of genes." Ref. Burks and Kelley ('28). See also Fisher ('18).
   c. The working out of an "exact science of mental functions, so that we could know the unite characters of mind as the biologist knows or expects to know the unit characters of plants and animals." Ref. Whipple's Manual p. 3. (NOTE: One of the psychologists interviewed for this report suggested that tests should be made which involve definite mechanisms and could be used on animals, thus enab'ing us to determine whether certain traits were unit characters or not. Ex. Dr. Yerkes' work on the inheritance of aggressiveness in the white rat, and the studies of Blatz and Burlingame on behavior traits in animals. See also Comparative Psychology,—studies of animal learning and the conditioned reflex.)

BIBLIOGRAPHY FOR APPENDIX II

Bere, M. '24. A comparative study of the mental capacity of children of foreign parentage. T. C. Contrib. to Educ., 154, Pp. 105.

Blackwood, B. '27. A study of mental testing in relation to anthropology. Ment. Meas. Monog., 4, Pp. 120.

Boody, B. M. '26. Testing immigrant children. Baltimore: Williams and Wilkins, Pp. 163.

Brigham, C. C. '26. Validity of tests in examinations of immigrants. Indus. Psychol., 1, 413-417.

—'30. Intelligence tests and immigrant groups. Psychol. Rev. 37, 158-165.

Burks, B. S. '29. Note on Prof. Freeman's discussion of the Stanford study of foster children. J. Educ. Psychol., 20, 98-100.

— and Kelley, T. L. '28. Statistical hazards in nature-nurture investigations. 27th Yrbk. Nat'l Soc. Stud. Educ., 8-38.

Colvin, S. S. '21. Intelligence and its measurement: a symposium. J. Educ. Psychol., 12, 136-139.

Dodd, S. C. '26. International group mental tests. Princeton, New Jersey: Princeton Univ. Store, agents. (Mimeographed) Pp. 101 plus appendix.

Fisher, R. A. '18. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal Soc. Edinburgh, 52, 399-430.

Freeman, F. N. '26. Mental tests: their history, principles and applications. Boston: Houghton Mifflin Co., Pp. 503.

Gates, A. I., and Taylor, G. A. '25. Experimental study of the nature of improvement resulting from practice in a mental function. J. Educ. Psychol. 16, 583-592.

Goodenough, F. L. '29. The emotional behavior of young children during mental tests. J. Juv. Res., 13, 204-219.

Haggerty, M. E. '21. Intelligence and its measurement: a symposium. J. Educ. Psychol., 12, 195-198.

Henmon, V. A. C. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 11, 195-198.

Kelley, T. L. '28. Crossroads in the mind of man. A study of differentiable mental abilities. Stanford University, Calif.: Stanf. Press, Pp. 238.

Kent, G. H., and Shakow, D. '28. Group tests for clinical studies. J. Gen. Psychol. 35, 595-618.

Klineberg, O. '28. An experimental study of speed and other factors in "Racial" differences. Arch. of Psychol., 93, Pp. 111.

Pintner, R. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 12, 139-143.

Pintner, R. '24. Results obtained with the non-language group test. J. Educ. Psychol. 15, 473-483.

Porteus, S. D. and Babcock, M. E. '26. Temperament and race. Boston: Badger, Pp. 364.

Rosenow, C. '17. The analysis of mental functions. Psychol. Rev. Monog. 24, Pp. 43.

Ruml, B. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol. 12, 143-144.

Shimberg, M. '29. An investigation into the validity of norms with special reference to urban and rural groups. Arch. of Psychol. 104, Pp. 84.

Squires, P. C. '26. A universal scale of individual performance tests. Examination Manual. Princeton, N. J., Prince. Univ. Press. Pp. 198.

Stockton, J. L. '22. Definition of intelligence in relation to modern methods of mental measurement. Psychol. Monog., 30, Pp. 118.

Terman, L. M. '21. Intelligence and its measurement. A symposium. J. Educ. Psychol., 12, 127-133.

—'24. The mental test as a psychological method. Psychol Rev. 31, 93-117.

Thorndike, E. L. et al '26. The measurement of intelligence. New York: Columbia Univ. Publ., Pp. 616.

Thurstone, L. L. '21. Intelligence and its measurement—a symposium. J. Educ. Psychol., 12 201-207.

Town, C. H. '16. Mental tests, a symposium. J. Educ. Psychol. 7, 352-353.

Travis, L. E. '28. The correlation between intelligence and speed in conduction of the nerve impulse in a reflex arc. Science, 67, 41-43.

Wallin, J. E. W. '16. Mental tests, a symposium. J. Educ. Psychol., 7, 353-357.

Watson, G. B. '26. Character tests of 1926. Vocatl. Guid. Mag. 5, No. 7, April.

Whipple, G. M. '22. Intelligence tests in colleges and universities. 21st Yrbk. Nat'l. Soc. Stud. Educ. Chap. 10, 262.

Whitely, M. T. '11. An empirical study of certain tests for individual differences. Arch. Psychol. 19, Pp. 146.

Wissler, C. '29. Reprint and circular series of the National Research Council. Final report of the Committee on scientific problems of human migration. Washington, D. C. 87.

Young, K. '24. The history of mental testing. Ped. Sem. 31, 1-48.

APPENDIX III

LIST OF THE CHIEF PERFORMANCE SCALES

Title: Army Performance Scale.
Devised by: Psychology Committee, National Research Council.
Date: 1920
Content:

Long Scale:

a. Ship Test.
b. Manikin and Feature Profile.
c. Cube Imitation.
d. Cube Construction.
e. Form Board 3 (Dearborn)
f. Memory for Designs.
g. Digit Symbol.
h. Maze Test.
i. Picture Arrangement.
j. Pictorial Completion II.

Short Scale:

b - d - f - g - h - j.

References:

Yoakum and Yerkes '20
Yerkes, (Ed.) '21.

Title: Arthur Performance Scale.
Devised by: G. Arthur.
Date: 1925
Content:

a. Cube Imitation.
b. Formboard (Seguin-Goddard-Norsworthy)
c. Two Figure Board.
d. Casuist Form Board.
e. Manikin and Feature Profile.
f. Mare and Foal.
g. Pictorial Completion I.
h. Block Design.

References:

Arthur '25 and '28.

Title: Boody Scale:
Devised by B. M. Boody.
Date: 1922
Content:

Group:

a. Beta.
b. Whipple's Cancellation (form h - l - p - g; "a" to be cancelled.)
c. Johns Hopkins Substitution Test. (Sym.-Dig.)
d. Johns Hopkins Number Test.
e. Substitution Test (Cincinnati Sym.-Dig.)
f. Memory Span for Digits. (Visual)
g. Dots (number in time limit).
N.B. e, f, g, were later discarded.

Individual:

h. Dearborn Formboard (No. 1-C).
i Card Sorting.
j. Johns Hopkins Coordination Test.

Stanford-Binet:

k. Divided Triangle.
l. Drawing a Square and Diamond.
m. Forms (matching).
n. Bow knot.
o. Drawings.
p. Estimating Weights.
N.B. k, l, m, n, o, and p were later discarded.

Later:

q. Army Picture Test (adapted Healy P. C. II).
r. Color-form-size Discrimination Test.

References:

Boody '26

Title: Scale used for Younger Children.

Date: 1923

Devised by B. M. Boody.

Contents: Stanford-Binet Revision.

a. Drawing of square.
b. Drawing of diamond.
c. Divided rectangle.
d. Bow knot.
e. Comparison of forms.
f. Drawing of cross, circle, star.
g. Button holes (2:4).
h. Pyramid with blocks (3:6).
i. Touching with fingers.
j. Discrimination test: color, form and size.

Reference:

Boody '26

Title: Condensed Non-Language Tests.

Devised by: Psychology Department, Boston Psychopathic Hospital. Edited by E. F. Symmes.

Date: 1924

Content:

a. Porteus Maze Tests.
b. Cube Immitation.
c. Adaptation Board.
d. Manikin Test.
e. Seguin Form Board (Goddard).
f. Mare and Foal Test.
g. Diagonal Test.
h. Substitution Test.
i. Triangle Test.
j. Picture Completion I.
k. Cube Construction.
l. Dearborn Formboard.
m. Construction Test A.
n. Construction Test B.
o. Picture Completion II.
p. Feature Profile.

Reference:

Mimeographed manuscript.

Title: Gaw Performance Scale.

Devised by F. A. Gaw.

Date: 1925

Contents:

a. Pictorial Completion I.
b. Pictorial Completion II.
c. Manikin.
d. Feature Profile.
e. Cube Construction.
f. Form Board. (Seguin, Goddard, Norsworthy).
g. Form Board 2. (Dearborn, Anderson).
h. Maze.
i. Cube Imitation.
p. Adaptation Board.
k. Substitution.
l. Triangle.
m. Diagonal.
n. Construction A.

Reference:

Gaw '25

Title: Knox Scale.

Devised by H. A. Knox.

Date: 1913

Content:

a. 65 tests (including repetitions) for ages 3 to 13 composed of many elements from Binet tests and Knox's own performance tests, etc.

References:

Knox '13
Knox '14

Title: Maturity Scale. (Mixed performance and verbal)
Devised by Dewey, Child and Ruml.
Date: 1920
Content:

a. Comprehending Questions.
b. Hard Definitions (Yerkes, Bridges, Hardwick).
c. Needle Threading (for girls alone).
d. Cancellation.
e. Automobile Construction.
f. Card Sorting.
g. Problem Box.
(e, f, g, for boys alone.)
h. Resisting Suggestions.
i. Dissected Sentences (Yerkes, Bridges, Hardwick).

References:

Dewey, Child, Ruml '20

Title: Performance Scale.
Devised by E. L. Cornell and Coxe.
Date: In press.
Content:

a. Manikin and Profile.
b. Block Designs (reproduce designs not completed).
c. Picture Arrangement.
d. Digit Symbol.
e. Memory for Designs (a, b, c, d, d2).
f. Cube Construction (a, b, c).
g. Picture Completion.

References:

To be published by Public School Publishing Company.

Title: Pintner, Paterson Performance Scale.
Devised by: R. Pintner and D. G. Paterson.
Date: 1917
Content:

Long Scale:

a. Mare and Foal.
b. Seguin Form Board.
c. Five Figure Board.
d. Two Figure Board.
e. Casuist Form Board.
f. Triangle Test.
g. Diagonal Test.
h. Construction Test A.
i. Manikin.
j. Feature Profile.
k. Ship Test.
l. Pictorial Completion Test I.
m. Substitution.
n. Adaptation Test.
o. Cube Imitation.

References:

Pintner-Paterson '17
Johnson '25

Title: Learning Tests.
Devised by W. H. Pyle.
Date: 1925
Content:

a. Card Sorting.
b. Marble Sorting.
c. Manthanometer.
d. Substitution (digit-symbol).
e. Mirror-ideational.
f. Mental Tests (Pyle's Missouri Tests).

References:

Pyle '25

Title: A Scale of Individual Tests (mixed).
Devised by Stevenson Smith.
Date: 1927
Content:

a. Color Naming.
b. Digit Recall.
c. Cancellation.
d. Association.
e. Picture Recall.
f. Perception of Number.

References:
S. Smith '27

Title: Simpson's Scale (mixed).
Devised by B. R. Simpson.
Date: 1912
Content:

a. Cancellation of a's.
b. Marking geometric forms in a scrambled series.
c. Scroll test—tracing a spiral pencil maze.
d. Recognizing forms test (pictorial).
e. Learning pairs; e. g., Chinese characters and English meaning.
f. Estimating lengths.

References:
Simpson '12

Title: Steacy Scale (mixed).
Devised by F. W. Steacy .
Date: 1919
Content:

a. A1 and A2. The Stenquist Construction Test. Single Series 1.
b. A3 and A4. The Stenquist Recognition of Mechanical Devices Test.
c. A5 and A6. A Knot Making Contest, Ruger, Mandl, Series I and Series II.
d. A7 and A8. Ruger Puzzle Series. I, II, III, IV.
e. A9 and A10. Diagram Matching Test. Series A, Series B.
f. A11 and A12. Diagram Drawing Test Series C. Series D. Spatial relations test.
g. B1 Superposition Test.
h. B2 Rogers Modification of Thurstone's Spatial Relations Test.
i. B5 Similar Figures, No. 17. Briggs.

References:
Steacy '19

Title: Steggerda Scale (mixed).
Devised by M. Steggerda.
Date: 1929
Content:

a. Seashore Musical Tests.
b. Copying three geometric figures.
c. Drawing a man.
d. Manikin.
e. Knox Moron.
f. Paper folding test (S-B-18 yrs.)
g. Memory span for numbers: 7 figures.
h. Binet absurdities.
i. Army Alpha.

References:
Davenport '29

Title: Universal Scale of Individual Performance Tests.
Devised by P. C. Squires.
Date: 1926
Content:

a. Series Forward.
b. Single Cube Piles.
c. Association Test (Pictorial)
d. Association Test (Quantitative)
e. Arrangement of Cube Piles Test.
f. Painted Cube Test.
g. Series Reversed Test.
h. Geometric Association Test.
i. Block Construction Test.
j. Picture Analysis Test.

k. Reasoning Test.
l. Free Analogies Test.
m. Spot Pattern Test.
n. Right and Left Hands Test.

References:

Squires '26

Title: Woolley Scale of Mental Ability. (mixed).
Devised by H. T. Woolley.
Date: 1913
Content:

a. Cancellation.
b. Substitution.
c. Digit Memory Span.
d. Completion of Sentences.
e. Opposites.
f. Mutilated Text.
g. Construction Puzzles.
h. Puzzle Box.
i. Instruction Box.
j. Visual Recognition.
k. Aussage.
l. Hard Directions.

References:

Woolley '13

NOTE: All references in Appendix III to be located in Bibliography of References on Tests, Part III, pages 234–242.

APPENDIX IV

LIST OF BIBLIOGRAPHIES ON INTELLIGENCE TESTS

Bell, J. C. '21. Group tests of intelligence, an annotated list. J. Educ. Psychol. 12, 103-108. 30 ref. (ann.)

Bibliography of tests for use in schools. World Book Co., pp. 36, price 10 cents. Contains 528 titles of tests, classified by subject, under the names of authors or titles. Gives grades for which intended and names of publishers.

Boardman, H. '18. Psychological tests, a bibliography. Bur. of Educ. Exper. Bull. No. 6, pp. 75.

Brigham, C. C. '17. The diagnositic value of some mental tests. Psychol. Monog. 24, No. 102, 95-254. 82 ref.

Bureau of Education. Dept. of Interior '23. Bull. No. 55. Compiled by Doherty M., MacLatchy J. under the direction of Buckingham, B. R. Washington: Govt. Print. Off.

Burr, E. '22. Psychological tests applied to factory workers. Arch. Psychol. No. 55, pp. 93, 60 ref.

Carothers, F. E. '22. Psychological examinations of college students. Arch. Psychol. No. 46 pp. 82, Bibl. fn.

Eliot, A. A. Psychological tests for children under six years of age. Unpublished. Report sent by author to Committee on Child Development, National Research Council.

Freeman, F. N. Tests.

Psychol.	Bull.	1911- 8- 21- 24
"	"	1912- 9-215-222
"	"	1913-10-271-374
"	"	1914-11-253-256
"	"	1915-12-187-188
"	"	1916-13-268-271
"	"	1917-14-245-249
"	"	1919-16-374-381
"	"	1920-17-353-362

Gambrill, B. L. '27. An analytical list of kindergarten-primary tests of intelligence and achievement. New Haven, Conn.: Whitlock's Book Store Inc..

Geyer, D. L. '26. Selected references on standardized tests and statistics. Chicago Sch. J. 8, 219-223. 103 ref.

Great Britain Board of Education. Consultative Committee. '24. Report on psychological tests of educable capacity and their possible use in the public system of education. London: 200 pp. 300 ref.

Jensen. D. W. '27. The gifted child. J. Educ. Res. 15:34-45. 126-133. 106 ref.

Jordan, A. M. '23. The validation of intelligence tests. J. Ed. Psychol. 14, 414-428. 62 ref.

Kelley, T. L. and Briggs, T. H., Tillinghast, C. C. and others. '22. Bibliography of tests for use in high schools. Teach. Coll. Rec. 23, 375-395.

Kohs, S. C. '14. The Binet-Simon measuring scale for intelligence,an annotated bibliography. J. Educ. Psychol. 5, 214-224, 279-290, 335-346, 254 ref.

Louttit, C. M. '28. A bibliography of bibliographies on psychology, 1900—1927. Nat. Res. Counc. Bull. No. 65, p. 108. Includes all the important bibliographies which have been published for the period. Good index.

Manson, G. E. '26. A bibliography of the analysis and measurement of human personality up to 1926. Nat. Res. Counc. and Cir. Serv. No. 72, pp. 59. 1364 ref.

Merrill, M. A. '24. On the relation of intelligence to achievement in the case of mentally retarded children. Comp. Psychol. Monog. 2, No. 10, pp. 100, 63 ref.

Mitchell, D. and Ruger, G. J. '18. Psychological tests. Revised and classified bibliography. New York. Bur. Educ. Exper. Bull. No. 9, pp. 116. 1428 ref.

Monroe, W. S. '20. A bibliography of standardized tests for the high school. J. Ed. Res. 1, 148-153, 228-242, 311-320. (ann.). Reprinted separately: Bloomington, Ill. P. S. Pub. Co.

Odell, C. W. '23. An annotated bibliography dealing with the classification and instruction of pupils to provide for individual differences. Univ. of Ill. Bull. No. 16, Urbana.

Pintner, R. Intelligence Tests. Psychol. Bulletin. 1926, 23, 366-381
1927, 24, 391-408
1928, 25, 389-406
1929, 26, 381-396
1930, 27, 431-457

Seago, D. W. '25. An analysis of the language factors in intelligence tests. Ment. Meas. Monog. No. 1, 124.

Whipple, G. M. '23. An annotated list of group intelligence tests. 21st Yrbk. N. S. S. E. Bloomington, Ill. 93-113.

Wylie, A. T. '20. A brief history of mental tests. Teach. Coll. Rec. 33, 19-33.

Young, H. H. '16. The Witmer formboard. Psychol. Clin. 10, 93-11. 12 ref.

PUBLISHERS AND MANUFACTURERS HANDLING TEST MATERIALS.

College Book Company, Columbus, Ohio.
Education and Personnel Publishing Company, Washington, D. C.
Eugenics Record Office, Cold Spring Harbor, New York.
General Electric Company, Lynn, Massachusetts.
Mabel Goudge, 102½ West Main Street, Durham, North Carolina.
Kansas State Teacher's College, Emporia, Kansas.
Marietta Apparatus Company, Marietta, Ohio.
Newson and Company, New York City.
Psycho-Educational Clinic, Palfrey House, Harvard University, Cambridge, Massachusetts.
Psychological Laboratory, Princeton University, Princeton, New Jersey.
Psychological Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania.
Public School Publishing Company, Bloomington, Illinois.
Research Service Company, 5259 S. Van Buren Place, Los Angeles, California.
C. H. Stoelting and Company, 424 No. Homan Avenue, Chicago, Illinois.
T. C. Bureau of Publications, Columbia University, New York City.
The Training School, Vineland, New Jersey.
Ellen Wilson, Rome, New York.
World Book Company, Yonkers-on-the-Hudson, New York.

INDEX